

令和4年度経済産業省委託
石油・ガス供給等に係る保安対策調査等事業

高圧ガス保安に係る
国家試験等の実施方法の在り方について
報告書

令和5年3月
高圧ガス保安協会

目次

要旨	1
1 事業の概要	1
1.1 課題	1
1.2 背景と目的	1
1.3 事業内容	1
(1) 国家試験等の実施方法の他事例調査	1
(2) 非 PBT 方式による国家試験等の在り方の検討	2
1.4 事業実施体制	2
1.5 調査方法	4
(1) 本事業における用語の定義	4
(2) 国家試験等の実施方法に係る他事例調査	4
2 試験理論概説	5
2.1 試験とは	5
2.2 試験が行われる背景	6
2.3 試験理論の必要性	7
2.4 試験理論	8
(1) 古典的試験理論 (CCT)	8
(2) 項目反応理論 (IRT)	10
(3) 最尤推定法	12
(4) Bayes 推定法	13
3 国家試験等の実施方法の他事例調査	14
3.1 アンケート調査 (試験実施団体)	14
(1) アンケート調査方法	14
(2) アンケート調査結果	16
3.2 ヒアリング調査 (試験実施団体)	26
(1) ヒアリング調査の概要	26
(2) ヒアリング調査の結果	26
3.3 非 PBT 事業者に関する整理・分析	31
(1) 非 PBT 事業者へのヒアリング調査の概要	31
(2) ヒアリング調査の結果	31
(3) PBT 方式と CBT 方式の試験に係る費用比較シミュレーション	36
4 非 PBT 方式による国家試験等の在り方の検討	39
4.1 非 PBT 方式で国家試験等を実施する場合の実施内容	39
(1) 試験の実施方式	39
(2) 出題方式	40
(3) 問題形式	41

(4) 問題数	43
(5) 出題問題の選定基準	43
(6) 問題評価	44
(7) 採点	45
(8) 合否判定基準	46
(9) 試験時間	46
(10) 試験回数（受験機会）	46
(11) 問題の公開	47
(12) 作問	49
(13) 受験会場	49
4. 2 非 PBT 方式で国家試験等を実施する場合の制度的制約	50
4. 3 国家試験等実施期間等の第三者による受験者本人の確認方法	54
4. 4 試験実施団体における国家試験等の非 PBT 化の有用性及び非 PBT 事業者の選定時のポイント	54
(1) 試験実施団体における国家試験等の非 P B T 化の有用性	54
(2) 非 PBT 事業者の選定	55
5 まとめ	60
付録書 A	67
付録書 B	97

要旨

本報告書は、経済産業省の令和4年度石油・ガス供給等に係る保安対策調査等事業「高圧ガス保安に係る国家試験等の実施方法の在り方について」に関する報告書である。

高圧ガス保安法及び関係省令で整備されている国家試験、講習検定（以下「国家試験等」という。）の目的のひとつは、特定の業務又は役職に従事又は選任される者に対してその法的、学術的、技術的な必要に応じた知識・技能の修得を公的に確認することによって、当該業務を望む者又は人材を選任する事業者の両者に「高圧ガスによる災害の防止」を前提とした高圧ガスに係る産業活動及び保安活動の水準維持を可能とすることにある。

本報告書では、第一に国家試験の実施方法の他事例調査を行うために、まず非 PBT¹方式に関する技術的な調査を行い、試験理論概説について整理した。続いて、資格試験（国家試験等を含む）を実施している機関・団体・学校他へ試験の実施状況に関するアンケート及びヒアリング調査の結果、非 PBT 方式の試験システムや試験運営サービスを提供する非 PBT 事業者へ試験システムの概要等に関するヒアリング調査の結果を取りまとめた。第二に、非 PBT 方式による国家試験等の在り方の検討を行うため、非 PBT 方式を実施する上でのポイント等について取りまとめた。

I 事業の概要

I. 1 課題

高圧ガス保安に係る国家試験等の実施方法の在り方

I. 2 背景と目的

今般の新型コロナウイルス感染症の感染拡大以降、多くの業界で実施していた会場集合型の試験等の実施方法の見直しが検討されている。高圧ガス保安協会においても、現在実施している高圧ガス保安分野に関する会場集合型の国家試験等の実施方法の在り方について、抜本的な検討が必要な時期に来ている。

本事業では、高圧ガス保安法以外の法令における国家試験や資格試験等における非 PBT 方式の試験の実施状況を調査した。また、CBT を念頭においた非 PBT 方式の試験に関する学術的・技術的な基礎調査及び市場の動向調査を行い、高圧ガス保安分野に関する国家試験等の実施方法の検討に資するよう取りまとめた。

I. 3 事業内容

(1) 国家試験等の実施方法の他事例調査

現に他法令で実施されている国家試験等（保安レベルを維持することを前提としたものに限る）の実施方法について、令和2年度経済産業省委託石油・ガス供給等に係る保安対策調査等事業（高圧ガス保安に係る法定講習の実施方法の在り方について報告書）の調査報告書も踏まえ、今回は非 PBT 方式で行われている国家試験等を中心に実態を調査した。

¹ Paper Based Testing。紙面（問題用紙と解答用紙）を用いた試験。

また、非 PBT 方式の試験制度の設計にあたっては、CAT (Computer Adaptive Test) 方式に代表される受験者の知識水準に応じた出題を行う、従来の PBT 方式における出題方法、評価基準を大きく変更する議論も必要であることから、非 PBT 方式に適した出題方法、運用方法に関する基礎調査を行い、高圧ガス保安法における国家試験の在り方の検討の一助となるよう整理し取りまとめた。

(2) 非 PBT 方式による国家試験等の在り方の検討

(1) で得られた調査結果を基に非 PBT 方式による試験の実施方法について、次の①及び②に基づいて検討を行い、非 PBT 方式による国家試験等の在り方をまとめた。

① 非 PBT 方式で国家試験等を実施する場合の実施内容及び制度的制約

非 PBT 方式で国家試験等を実施する資格制度としての連続性や法における制度趣旨が害されないかを確認する必要がある。したがって、高圧ガス保安法における従来の PBT 方式における出題方法又は評価基準との比較評価の議論を前提として、非 PBT 方式における標準的な国家試験等の実施方法について検討し、具体的な提案を行うこと。

② 国家試験等実施機関等の第三者による受験者本人の確認方法

高圧ガス保安に係る国家試験等について、現に受験者本人が受験していることの確認について、具体的な方法を検討し、実施案をまとめること。

1. 4 事業実施体制

1. 3 (1) 及び (2) までの調査を実施するにあたって、学識者、法定講習等実施機関、民間講習機関、非集合型教育業界団体の関係者により構成された委員会を設置し、高圧ガス業界団体をオブザーバに招くなどして、意見の聴取等を含めた事業の進行を図り、議論を取りまとめた。

高圧ガス保安に係る国家試験等の実施方法の在り方検討委員会 委員構成

(五十音順、敬称略)

- (委員長) 澁谷 忠弘 横浜国立大学 先端科学高等研究院
リスク共生社会創造センター センター長 教授
- 杉浦真由美 北海道大学 大学院教育推進機構オープンエデュケーションセンター
特任准教授
- 仲林 清 特定非営利活動法人 日本イーラーニングコンソシアム 副会長
千葉工業大学 情報科学部 教授
- 深田 章吾 一般財団法人日本ガス機器検査協会 教育講習部長
ガス主任技術者試験センター 所長
- 牧野 常夫 一般財団法人 全日本情報学習振興協会
理事長
- 宮坂 一弘 一般財団法人電気技術者試験センター
工事士試験部長
- 渡辺 明 株式会社東京リーガルマインド 資格事業本部
申請事業部 執行役員 第一種電気工事士定期講習担当

(事務局)

- 及川 裕幸 高圧ガス保安協会業務管理グループマネージャー
- 鳥越 利之 高圧ガス保安協会試験・検定作問グループマネージャー
- 高橋 智 高圧ガス保安協会業務管理グループ運営チームリーダー
- 木村 勝之 高圧ガス保安協会業務管理グループオンライン講習推進チームリーダー
- 高橋 元樹 高圧ガス保安協会試験・検定作問グループ作問チームリーダー
- 山崎 浩 高圧ガス保安協会試験・検定作問グループ作問チーム調査役
- 菊川 重紀 高圧ガス保安協会試験・検定作問グループ作問チーム調査役
- 篠田 康則 高圧ガス保安協会教育試験・検定作問グループ作問チーム
- 小田喜 洋介 高圧ガス保安協会業務管理グループオンライン講習推進チーム
- 海老原 塁 高圧ガス保安協会業務管理グループオンライン講習推進チーム
- 柴野 祐太 高圧ガス保安協会業務管理グループオンライン講習推進チーム
- 相良 尚都 高圧ガス保安協会業務管理グループオンライン講習推進チーム

審議経過

高圧ガス保安に係る国家試験等の実施方法の在り方検討委員会

第1回 令和4年 9月29日 事業実施計画について

第2回 令和4年12月19日 事業の進捗に関する報告

第3回 令和5年 3月16日 事業の報告

なお、審議は社会情勢に配慮して全てweb会議で実施した。

1. 5 調査方法

(1) 本事業における用語の定義

本事業において調査を行うにあたり、試験の実施方式である PBT (Paper-Based Testing)、CBT (Computer-Based Testing) 及び IBT (Internet-Based Testing) について、表 1 のとおり定義した。

表 1 試験実施方式の定義

実施方式		特徴
PBT 方式		方法：紙面（問題・回答用紙等）を配布・回収する方法 場所：指定された会場 時間：全国一斉に実施
非 PBT 方式	CBT 方式	方法：コンピューター端末（会場に設置のもの）を使用する方法 場所：指定された会場（テストセンター） 時間：①個別の時間帯で実施 又は ②全国一斉に実施
	IBT 方式	方法：コンピューター端末（任意のもの）を使用する方法 場所：任意の場所 時間：①個別の時間帯で実施 又は ②全国一斉に実施
その他		上記に分類されない方法 例：自宅学習型（郵送制度を利用し、回答、レポート提出等を行わせる方法）、実技試験

(2) 国家試験等の実施方法に係る他事例調査

高圧ガス保安法以外の法令における国家試験等及び民間の資格試験（以下「民間試験」という。）の実施方法に関する事例等を収集するために、アンケート調査（一次調査）を実施した。

アンケート調査の対象については、高圧ガス保安協会が抽出した法令等に基づく国家試験等の実施団体（42 ヶ所）と修了試験を伴う法定講習等の実施団体（10 ヶ所）に、非 PBT 方式の資格試験等を実施している民間団体等（15 ヶ所）を加えた計 67 ヶ所の試験実施団体とした。アンケート調査票を作成し、これらの試験実施団体に郵送した。アンケート調査の実施概要については、3. 1 を参照のこと。

上記アンケート調査の結果を踏まえ、アンケート調査に回答した試験実施団体のうち 10 ヶ所に対して CBT/IBT 化に関するヒアリング調査（二次調査）を実施した（3. 2 参照）。

また、資料調査の結果を踏まえ、非 PBT 方式の試験を実施するためのシステムやサービスを提供している事業者（以下「非 PBT 事業者」という）に対するヒアリング調査（二次調査）を実施した（3. 3 参照）。

2 試験理論概説

国家試験とは、国または地方公共団体あるいはそれらの委託を受けた団体が行う試験のことを指す。高圧ガス保安法においては、高圧ガスによる災害の発生の防止に関して、それぞれ必要な知識及び技能について国家試験を実施し、これに合格することで資格を取得することができる制度がある。また、国家試験以外の資格として、講習受講の後、技術検定試験に合格することで資格を取得する制度もある。

現在のPBT方式の試験の場合、今般の新型コロナウイルスの感染拡大を機に、受験者が急遽感染や濃厚接触者と判断され年に一度の受験機会を失う状況が生じる他、大規模な会場確保や採点等の試験事務の負担が生じる等の課題があることが強く認識された。このような状況を受け、非PBT方式の試験を視野に入れ、他法令の試験の実施状況とも比較しつつ、望ましい試験の在り方について検討を行った。議論の前提として、試験理論の概要を整理した。

参考として付録書Aを添付したため参照されたい。

2. 1 試験とは

試験とは、ある項目についての個人差を測定する手段として理解されている²。最も古い試験の形式は口頭質疑と考えられるが、有形の試験形式としては、古くは中国で1300年以上も人材選抜のために運用された科举、現代においては、学校教育で実施される学力試験などがその一例となる。試験は、その性質に注目すると、合格・不合格を判定する合否判定型の試験と能力の到達度を測定する能力測定型の試験の2つに分類でき、先の例では、科举は前者、学力試験は後者に当たる。

試験は1問以上の問題の集合で構成されており、各問題への解答を総合して評価するが、この評価には受験者本来の能力に加えて様々な誤差の影響があると考えられている。この誤差を大きく2つに分類すると、受験者個人の能力の振れ幅である系統誤差³と、試験の採点時等に生じる揺れ幅である偶然誤差⁴がある⁵。試験はこれらの要素を含み、大枠では2つの指標で評価される。1

² 『テスト（検査）は、人が持つ能力、学力、性格、適性、技能、関心、態度といった様々な諸相が個人ごとに異なること、つまり‘個人差’に着目し、その違いを量的に表現し具体化する方法として発達してきた。』池田央[2012]“テストの過去、現在、そして未来の形を考える”日本テスト学会誌、3頁

³ 『系統誤差は、同様の測定の繰り返しにおいて、各人ごとに一貫して生じる誤差であり、古典的テスト理論モデルでは真値に含まれるものである。』南風原朝和[2012]”尺度の作成・使用と妥当性の検討”教育心理学年報, Vol. 51, 214 頁

すなわち、ある試験における身体的、心理的条件による能力以外の要因による得点への影響。例として、問題・解答形式への慣れ、気温、机の状態、その日の体調等

⁴ 『偶然誤差は、同様の測定の繰り返しにおいて、その都度変動する誤差であり、古典的テスト理論モデルにおいて測定誤差とされるものである』南風原朝和[2012]”尺度の作成・使用と妥当性の検討”教育心理学年報, Vol. 51, 214 頁

すなわち、瞬間的な体調、環境の変化若しくは採点側の理由で発生する影響。例、瞬間的な頭痛、採点者間の採点基準の揺れ、採点者自身の心理的要因による揺れ等

⁵ 南風原朝和[2012]”尺度の作成・使用と妥当性の検討”教育心理学年報, Vol. 51, 213 頁

つは受験者の測定したい能力を測定できているかという妥当性⁶であり、もう1つは本人の能力を安定的に測定できているかという信頼性⁷である。妥当性が低いとは、測定したい能力を測れていないことを示し、信頼性が低いとは、測定結果が安定しないことを示す。信頼性を向上させるためには、偶然誤差を排除する必要がある、歴史的には採点誤差を生じる記述問題を客観的な採点ができる選択式問題に変更する方法が選ばれてきた。一方で、このような変更は、本来記述式問題で測定していた能力が測定できなくなることで妥当性が低下する可能性や、問題形式の変更により系統誤差が増大することで信頼性が低下する可能性があるため、慎重な検討を要する。

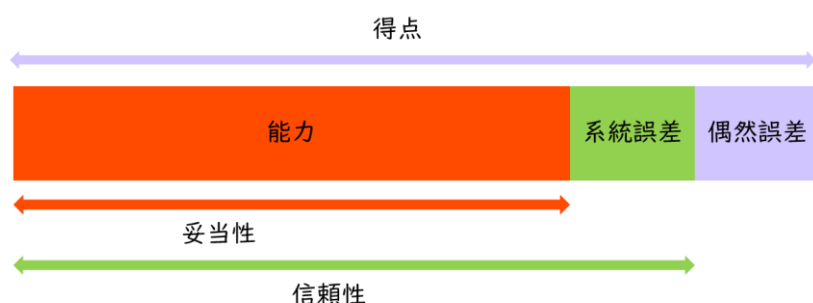


図1 得点の内訳

2.2 試験が行われる背景

現在、社会においてはさまざまな試験が活用されているが、それらは大きく分けて以下の役割をもつ場合が多い。

- 知能など、人間の心理的なある側面を測るテスト。語学能力やコミュニケーション能力のような複雑な概念を含む。
- 大学入試、就職試験などの「選抜」を目的とした試験。
- 資格試験に代表される、受験者の能力が一定の基準を満たしているか否かを検査するための試験。

これらの試験は測定される対象に応じて、さまざまな実施形態で行われている。自動車運転免許を取得する際には、実際に公道を運転する前に教官が教習所にて運転技能を見極めることが行われるが、これも広義の「試験」にあたる。また英語の試験では、面接者と受験者が口頭で英語のやり取りをし、その過程を面接者や評価者が評定するという形をとる場合がある。これらのように、人間のパフォーマンス（運転能力、英語会話能力など）を実際に観測し、それらを評価者が評定するタイプの試験は「パフォーマンステスト」と呼ばれている。

それに対し、思考力や判断力、知識といった内容を測るために、多数の問題項目（以下、受験者に提示する問題項目のことを「項目」と表記する）を受験者に提示し、解答を収集して採点し、得点（スコア）を算出するタイプの試験もある。このタイプの試験はいわゆる「紙と鉛筆」による試験の場合もあれば、コンピュータを用いて項目を提示し、解答もコンピュータで収集するも

⁶ 測定対象の能力を測定できているか。内容（試験範囲）的な視点と認知（難易度等）的な視点がある。

⁷ 結果が状況に左右されずに一貫して得られるか。内部一貫法、再テスト法や平行テスト法、折半法等で評価。相関係数 ρ 。信頼性係数 $\rho^2 = V(T)/V(X)$ or $\rho^2 = 1 - V(E)/V(X)$ V は分散、 T は能力、 X は得点、 E は誤差。

のもある。また「記述式」や「多肢選択式」といったさまざまな問題形式で行われるが、一般にこのタイプの試験が、世間一般で「試験」と聞いてイメージするものであろう。

試験実施の際に使用される各用語については、次のように定義している。

測定：ある一つのものごとについて、何らかの程度の大きさを数値化し、表示しようとする
こと。測定にあたっては、何らかの道具（たとえば長さの測定にあたっては「ものさし」、英語能力の測定のためには「英語の試験」）を用意し、定められた手続きによって行われる。

採点：試験を用いた人間の能力の測定において、項目に対する正誤を判断し、受験者ごとに得点を求める手続き。パフォーマンステストでは次項で示す「評価」を行い、その数値をもって得点を求める手続きを行うが、ここでは「採点」に含めない。

評価：パフォーマンステストにおいて、受験者ごとにパフォーマンスを見定め、あらかじめ定められた基準にもとづきその程度の大きさを数値化すること。記述式試験においても「記述された答案」をパフォーマンスの一種にとらえ、パフォーマンステストと同様の操作を行うが、記述式試験の場合は「採点」と表記し、区別している。

評価：人間の能力の大きさを、採点や評価に基づき数値化・得点化・カテゴリ化し、その傾向の大きさを示すこと。評価の結果は数字で示される場合もあれば、「合格／不合格」「優・良・可・不可」のようにカテゴリで示される場合もある。

2. 3 試験理論の必要性

各試験形式について、試験実施から採点、得点の算出までの流れを記すと、以下のようになる。

● パフォーマンステスト

- ① 受験者からパフォーマンスのデータ（実技のデータなど）を収集する
↓
- ② 評価者があらかじめ定められた評価基準に基づき、評価点をつける。
↓
- ③ 評価点を集計し、受験者ごとに得点を算出する。

● 記述式試験

- ① 受験者から記述式答案を集める。
↓
- ② 採点者があらかじめ定められた採点基準に基づき、採点する。
↓
- ③ 採点結果を集計し、受験者ごとに得点を算出する。

● 多肢選択式試験

- ① 受験者から解答データ（項目ごとに、どの選択肢を選んだかを記録したデータ）を収集する。

↓

② 解答データを正答キー（正答選択肢）と突き合わせ、採点する。

↓

③ 必要に応じて配点をかけ、受験者ごとの得点を算出する。

パフォーマンステストと記述式試験は、評価者（採点者）が評価（採点）することをはじめ、その多くのプロセスが共通した要素を含むことがここからわかる。一方で、多肢選択式試験は評価者（採点者）がいらないため、より客観的な採点が可能であることから「客観式試験」と呼ばれることがある。ただし、出題内容や選択肢の構成によっては「正解」「不正解」が明確に分かれない場合があるため、多肢選択式試験であっても出題内容の吟味を重ねる必要性は残る。

いずれの試験形式とするかを問わず、試験は「人間の能力の一面面を測定するための尺度（ものさし）を構成するためのもの」ということができる。英語の試験は「人間の英語能力を測定するためのもの」であり、出題される問題項目は「人間の英語能力を測定するための道具」と位置付けられる。

ここで述べた3種の形式の試験のうちどの形式をとるにしても、採点結果（得点）が、受験者が実際にもっている能力を正確に反映した得点を返すことが求められている。評価者（採点者）が評価する試験の場合は、その評価者の質が一定水準に満たないと、受験者の能力を正確に反映した評価とはならなくなるおそれがある。また多肢選択式試験の場合は、項目の質が悪いと、受験者の能力とは関係ない要因で正解・不正解が分かれることとなってしまう、試験の質が悪化する原因となってしまう。

「試験の質」の良し悪しを評価するためには、出題された項目の内容的分析だけではなく、受験者の解答行動や評価者の評価をもとにした量的データに基づく分析が行われる。この量的データの分析を行う枠組みとして提案されたのが「試験理論」である。

2. 4 試験理論

（1）古典的試験理論（CCT）⁸

古典的試験理論は、試験で正解した問題数と対応する配点を計算することで受験者の能力を得点という形で評価する試験理論である。しかし、この得点という値は受験者の能力を示す指標としては互換性に乏しく、ある試験及び配点条件における結果でしかないため、異なる試験を受験した人との得点の比較をすることができない⁹。この課題は偏差値¹⁰等の指標を用いることで解決することができるが、この場合においても受験者が同一でなければならないという制約がある。すなわち、偏差値等は学校等で受験者が固定されている場合には、集団が受験した異なる試験間の結果を比較することができる有効な手段であるが、別の学校で行われた別の試験の結果との比

⁸ Classical Test Theory : CCT。

⁹ 例えば1問1点の問題の試験と1000点の問題の試験での得点比較、同じ配点だが問題が異なる試験での得点比較（小1レベルと中1レベルの試験比較）は試験内容や両者の関係性を示す指標がないと比較ができない。

¹⁰ 偏差値とは、標準得点（後述）を10倍して50を加えたもの。平均値は50、標準偏差は10になる。集団内での相対的な位置を示すものであるため、母集団が変わると比較ができない。偏差値60以上の人は、受験者全体の上位約16%に相当し、偏差値70以上の人は、受験者全体の上位約2.3%に相当する。

較は行えないということになる。

上述の試験結果の比較の例として、以下の表 2 に示す受験者 A から F までの 6 人が受験した 3 回の試験の結果を仮定して検討する。3 つの試験でそれぞれ 90 点を取った受験生 B、C、E の能力の高低を考える場合、比較の方法として直感的な思考手順に沿った以下 3 パターンの例を挙げる。

表 2 試験結果

	試験 1	試験 2	試験 3
受験者 A	59	9	39
受験者 B	90	36	40
受験者 C	78	90	34
受験者 D	49	15	14
受験者 E	41	66	90
受験者 F	97	48	47
平均	69	44	44
標準偏差	20.8	28.1	23.0

- ① 3 人の 90 点という得点に注目し、受験者 B、C、E を同じ能力と評価する
- ② 試験 1 と試験 2、試験 3 では試験の平均点が異なり、試験 2、3 は平均点の低い難しい試験であると考えられることから、受験者 B より受験者 C、E の能力の方が高い
- ③ 試験 2、3 では試験 3 の方が標準偏差¹¹が小さく、点が重い試験であるため受験者 C より受験者 E の能力の方が高い。

①は試験の難易度を考慮しておらず、②は点の取り難さを考慮していないため、正しく比較できていない。③のように標準偏差を考慮すれば今回のように点数が同じ受験者の比較は容易にできるが、受験者同士の点数が異なる場合や大規模試験を想定した場合には、この方法での比較は困難である。この課題については、全体を標準得点¹²に変換する標準化処理を施すことで解決することができる。表 3 は表 2 に対して標準化処理を行った結果であり、表 4 は標準得点と同じ概念である偏差値に変換したものである。試験結果の最大値で比較をする場合には、試験 3 における受験者 E の結果が最も偏差値が高く、受験者 E がこの集団のなかでは最も高い能力を発揮したことがあるということが分かる。

ただし、前述の通り、標準得点や偏差値による比較は、ある集団内における相対的な位置を表すものであるため、受験した集団が異なる場合には比較することができない。この課題を解決す

¹¹ 標準偏差とは、分散（後述）の平方根のこと。本質は分散と同じであるが、分散は、元となる平方和（後述）が 2 乗で求められていることから単位が 2 乗されており、標準偏差はこれを解決しているという意味がある。なお、分散は平方和の欠点を解消したもので平方和/データ数で求められ、平方和は平均からの振れとして（個々のデータ-平均）の 2 乗を合計したものと計算される。平方和はデータの個数が多くなるほど値も大きくなるという致命的な欠点があるためあまり用いられない。

¹² （個々のデータ-平均）/標準偏差で算出する。基準値ともいう。標準得点の平均値は 0、標準偏差は 1 になる。集団内での相対的な位置を示すものであるため、母集団が変わると比較できない。

るためには、集団間で共通の問題を解答させて集団の能力差に対する相関を導く必要がある。しかし、同じ問題を同じ条件で解かせることは、予め比較を想定していた集団にしか適用できず、事後的に比較を行うためには、間接的に相関係数を紡いでいくような手法をとることとなり運用が難しい。

表 3 試験結果の標準得点

	試験 1	試験 2	試験 3
受験者 A	-0.48	-1.24	-0.22
受験者 B	1.01	-0.28	-0.17
受験者 C	0.43	1.64	-0.44
受験者 D	-0.96	-1.03	-1.31
受験者 E	-1.35	0.78	2.00
受験者 F	1.35	0.14	0.13
平均	0	0	0
標準偏差	1	1	1

表 4 試験結果の偏差値

	試験 1	試験 2	試験 3
受験者 A	45.2	37.6	47.8
受験者 B	60.1	47.2	48.3
受験者 C	54.3	66.4	45.6
受験者 D	40.4	39.7	36.9
受験者 E	36.5	57.8	70.0
受験者 F	63.5	51.4	51.3
平均	50	50	50
標準偏差	10	10	10

(2) 項目反応理論 (IRT) ¹³

項目反応理論とは、受験者の能力を確率的に求めるという試験理論である。古典的試験理論と大きく異なる点は、出題する問題の評価を事前に行うという点であり、これにより問題に対する受験者の解答を能力推定の判断材料とすることができるようになる仕組みである。また、評価済みの問題を用いて試験を行うことから、試験結果は、いつ、どこで、どんな問題セットの試験を受けたのかということに影響されることなく、常に他と比較可能かつ数学的な根拠を伴ったものとなる。

¹³ Item Response Theory、IRT。項目とは試験における問題のことを指す。

	問題 1	問題 2	問題 3	問題 4	問題 5	問題 6	
受験者 A	○	×	×	×	×	×	
受験者 B	○	○	×	×	×	×	
受験者 C	○	○	○	×	×	×	
受験者 D	○	○	○	○	×	×	
受験者 E	○	○	○	○	○	×	
受験者 F	○	○	○	○	○	○	← 優秀
						↑ 難問	

図 2 試験に対する評価の視点のイメージ

図 2 は、試験に対する評価視点のイメージであるが、試験に対する評価の視点は受講者と問題の 2 つの側面がある。古典的試験理論では、試験後に横の行を評価することで 6 問中 6 問正解できた受験者 F は優秀であるとし、縦の列を評価し問題 6 は正答率が 17%であったため難問であったという評価をしている。ここで、6 問中 5 問正解した場合、本来どの問題に誤答したかによって、受験者の能力に差異があるはずであるが、古典的試験理論ではこれを測ることが難しい。これは配点という段階的な評価をしていることで情報を単純化していることに加えて、どの問題がどんな能力の識別にどの程度有効であるかの情報が無いためである。

一方、項目反応理論では、予めどの問題がどんな能力の識別にどの程度有効であるかという情報を連続的な確率分布という形で取得しているため、6 問中 5 問正解の受験者の解答パターンについても能力に差をつけて推定することができる。すなわち項目反応理論は、試験前に縦の列に関する評価を行い、試験後に横の行の評価を行うという仕組みで運用されるものといえる。

視力検査を例にとった場合、検査前のランドルト環の評価から図 3 のようなグラフが得られる。この図が意味するところは、このランドルト環に対して視力 0.46 以下の者は全員間違った解答をし、視力 0.47 から 0.51 の者は正解したり、誤答したりし、視力 0.52 以上の人は全員正しい解答をするということである。すなわち、このランドルト環は視力 0.5 を判別する問題として適性があるという評価になる。視力検査後は、このランドルト環に対する解答結果（及びその他の問題の結果）を基に受検者の視力を推定するという流れになる。

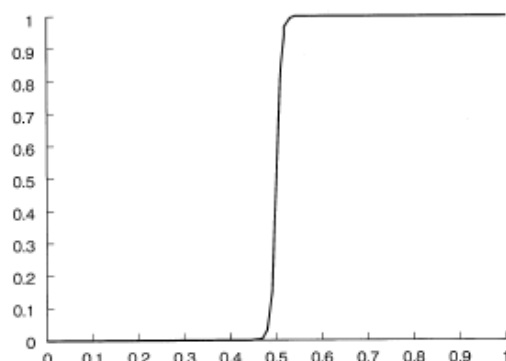


図 3 視力検査の項目反応関数¹⁴のグラフ（横軸：視力、縦軸：正解確率）

¹⁴ 光永悠彦[2022]”テストはなんのためにあるのか 項目反応理論から入試制度を考える”, ナカニシヤ出版, 91 頁

このように、項目反応理論では、受験者の能力として尤もらしい確率を推定するが、この推定の過程は、②解答結果がある、①この解答をする受験者の能力はこうである、というものとなり、実際の時間軸（①ある能力を持った受験者がいる、②この受験者が解答するとこの結果になる）と逆行している。このような時間逆行の確率を求めることは数学的に難しい処理となり、これを簡便に行うためにいくつかの手法が存在している。以下の図4は時間逆行の確率を求めることの難しさを示した例題となる。

（例題）

袋大（赤6、白4）、袋小（赤1、白4）を前提にして各確率を求める。

$$P(\text{大}) = \frac{1}{2} \quad \leftarrow \text{袋大を選ぶ確率}$$

$$P(\text{赤}) = \frac{1}{2} \times \frac{6}{10} + \frac{1}{2} \times \frac{1}{5} = \frac{2}{5} \quad \leftarrow \text{赤を選ぶ確率}$$

$$P(\text{大} \cap \text{赤}) = \frac{1}{2} \times \frac{6}{10} = \frac{3}{10} \quad \leftarrow \text{袋大の赤を選ぶ確率}$$

$$P(\text{赤} | \text{大}) = \frac{6}{10} = \frac{3}{5} \quad \leftarrow \text{袋大から赤を選ぶ確率}$$

$$P(\text{大} | \text{赤}) = \frac{P(\text{大} \cap \text{赤})}{P(\text{赤})} = \frac{\frac{3}{10}}{\frac{2}{5}} = \frac{3}{4} \quad \leftarrow \begin{array}{l} \text{選んだ赤が袋大のものである確率} \\ \text{これだけ時間逆行で難しい} \end{array}$$

図4 IRTにおける受験者能力の推定のイメージ

項目反応理論を用いた試験の代表的な例には、TOEIC¹⁵やITパスポート試験がある。

（3）最尤推定法¹⁶

最尤（さいゆう）推定法とは、ある現象を起こした対象の能力（現象を起こす確率）がどの程度のものであったのかを推定する方法である。最尤推定法の特徴は仮定にあり、本来、固定値である能力を変数と仮定するところから始まる。その仮定のもとで現象が起こる確率（尤度）を求め、そこから逆算的に仮定した能力を推定していく方法をとる。すなわち、ある現象の起こりやすさが最も高くなる時の能力が、現象を引き起こす対象の能力として尤もらしい¹⁷という推定である。ここで、最尤推定法で求まる結果について端的に表現すると、これは現象が起こる確率

¹⁵ Test of English for International Communication。国際コミュニケーション英語能力テスト。試験の結果は10-990点のスコアで示される。スコアの特徴として、全問不正解でも10点を取れ、数問間違えても満点を取れる仕組みとなっており、これはIRTで推定した能力値をスコアに換算するという処理を行っているためである。

¹⁶ Maximum Likelihood Estimation, MLE

¹⁷ 最尤推定法で求めた推定値は最尤推定値と呼ぶ。

と同じ値を返している。具体的には以下の例題を見ていただくと分かりやすい。

(例題)

コインを3回投げて2回表であったという事実に対して最尤推定法でこのコインの表を出す能力を推定すると、 $\frac{2}{3}$ (66.7%)となる¹⁸。しかし、この推定は我々の事前知識(コインが表を出す確率は $\frac{1}{2}$ (50%))と乖離している。

このように現象が起こった確率をそのまま対象の能力とする結果は、対象の能力発揮の回数が少ない場合には、1回の誤差の影響が大きく反映されてしまう。最尤推定法の問題点は、試行回数が少ない現象に対しては対象の能力推定の妥当性が低くなることにある。また、全問正解、全問不正解といった絶対値が大きい結果に対しての推定も行えないという欠点がある。

(4) Bayes 推定法

Bayes 推定法は、最尤推定法と同様にある現象を起こした対象の能力を求める方法である。Bayes 推定法の特徴は、対象が一般的な集団に属しているものと考えて対象の一般的な能力を事前知識(理論値)として仮定する点にある。この事前知識を基に現象が起こる確率を求め、これが最大になる値を能力として推定している。こうして得られる結果について端的に表現すると、現象が起こった確率と事前知識の能力の中間帯の値が結果として求まる。

(例題)

コインを3回投げて2回表であったという事実に対して Bayes 推定法を用いる際、事前知識として beta 分布で $C \times p^{50}(1-p)^{50}$ (100 回投げて 50 回表が出る。標準偏差は 5%) を仮定した場合、このコインの表を出す能力を推定すると、およそ 50.5%になる¹⁹。この値は、我々の事前知識であるコインの表を出す能力が 50%であることとの乖離が小さく、現象(表が多く出ている事実)も反映しているそれらしい値であると考えられる。Bayes 推定法では、試行回数が小さいときは事前知識の値(理論値)に近づき、試行回数が大きいときは逆に事前知識の影響が小さくなり、対象の真の能力を表すことになるため、最尤推定法で欠点とされていた試行回数が少ない場合の推定も解決しており、数学的な根拠もよりしっかりとした推定となる。ただし、Bayes 推定法においても、事前知識を仮定するという部分に自由度があり、結果に大きな影響を与えるため注意が必要である。

¹⁸ コインの表を出す能力を p と仮定すると、コインを3回投げて2回表が出た事象は ${}_3C_2 p^2(1-p)$ の確率で起こる。この確率が最大になるときの p は、 ${}_3C_2 p^2(1-p)$ の対数関数(簡便のため)

$2 \log p + \log(1-p) + \text{Const}$ の導関数 $\frac{2}{p} - \frac{1}{1-p}$ が0になる場合であるため、 $\hat{p} = \frac{2}{3}$

¹⁹ コインの一般的な能力 ${}_{100}C_{50} p^{50}(1-p)^{50}$ を前提として、コインを3回投げて2回表である確率は、 ${}_{100}C_{50} p^{50}(1-p)^{50} \times {}_3C_2 p^2(1-p)$ であり、これは $C \times p^{52}(1-p)^{51}$ であるので、最尤推定法と同じように計算をすると $\hat{p} = \frac{52}{52+51} \doteq 50.49\%$ となる。

3 国家試験等の実施方法の他事例調査

3. 1 アンケート調査（試験実施団体）

国家試験等の在り方を検討するためには、他法令において同様に位置づけられる資格試験に関する実施方法の把握が重要であることから、資格試験を実施している機関・団体・学校に対し、「国家試験等の実施状況に関するアンケート調査」（以下「アンケート調査」という）を実施し、取りまとめた。

（1） アンケート調査方法

総務省行政評価局の平成 23 年度の資料にある「国の資格制度一覧」（313 資格）²⁰及び高圧ガス保安協会の令和 2 年度の調査報告書の資料にある「アンケート調査先」²¹を基に、国家試験等の実施機関・団体・学校及び民間試験の実施機関・事業者（以下「試験実施団体」という）へのアンケート調査や必要に応じてヒアリング調査を行い、高圧ガス保安法以外の法令における国家試験等の実施方法について調査するとともに、保安レベルを維持する国家試験等の実施方法として有効な手段を検討した。

試験実施団体へ非 PBT 方式の国家試験等の実施有無に関するアンケートを実施した。アンケート調査の概要については、表 5 のとおりである。アンケート調査票²²は後述する。

表 5 アンケート調査の概要

調査対象	国家試験等を実施している機関・団体・学校
調査依頼先数	67ヶ所 (内訳：国家試験(講習なし)42ヶ所、国家試験(講習あり)10ヶ所、 民間試験(講習なし)15ヶ所)
調査期間	2022年12月26日～2023年1月18日
回答方法	① 郵送（アンケート調査票に記入し返送） ② Web（Google フォームに回答）

上記調査先へ送付したアンケート調査票の質問や選択肢等を整理・要約したものを、以下の表 6 に示した。

本調査について、大きく分けて「資格試験の実施状況について」と「非 PBT 方式の試験について」に関する質問を設定した。「資格試験の実施状況について」に属する質問は全ての試験実施団体を対象とし、1つの試験実施団体において複数の試験を実施していることがあるため、1試験実施団体につき代表的な試験2つまで回答可とした。「非 PBT 方式の試験について」に属する質問は、基本的に非 PBT 方式の試験実施団体が対象だが、「不採用の理由」は PBT 方式の試験実施団体、「提供できる情報」は全ての試験実施団体を対象とした質問とした。

²⁰ https://www.soumu.go.jp/main_content/000131811.pdf 資-130 頁

²¹ https://www.meti.go.jp/meti_lib/report/2020FY/000355.pdf 資-127 頁

²² 付録書 B アンケート調査票

表 6 アンケート調査の質問項目等

【Q 1】機関・団体・学校名

【Q 2】担当者連絡先

【Q 3】資格試験の実施状況

質問項目	回答形式	選択肢
試験名称	記述式	
問題の公開	選択式 (複数回答)	・ 常に HP 等で公開 ・ 一定期間のみ HP 等で公開 ・ 過去問題集を発刊 ・ 公開していない ・ その他※
本人確認の為の 事前提出物	選択式 (複数回答)	・ 受験票 (写真有) ・ 受験票 (写真無) ・ 公的身分証 (写真有) ・ 公的身分証 (写真無) ・ 免状や資格証 (写し) ・ 特になし ・ その他※
直近の年間受験者数	記述式	
問題の解答形式	選択式 (複数回答)	・ 選択式 (択一・択多) ・ 記述 ・ 口述 ・ 実技 ・ その他※
現在の試験実施方式	選択式 (複数回答)	・ PBT 方式 ・ CBT 方式 ・ IBT 方式 ・ その他※
今後の試験実施計画	選択式 (複数回答)	・ 非 PBT 方式の試験を中心に実施予定 ・ PBT 方式と非 PBT 方式の試験の両方を実施予定 ・ PBT 方式の試験を中心に実施予定 ・ その他※

- 非 PBT 方式の試験実施について

【Q 4】試験システムについて

質問項目	回答形式	選択肢
試験システム	導入方法	選択式 (単一回答)
	外部委託した業者 名・サービス名	記述式

【Q 5】試験会場の運営方法について

質問項目	回答形式	選択肢
試験会場	運営方法	選択式 (単一回答)
	外部委託した業者 名・サービス名	記述式

【Q 6】採用の理由

回答形式	選択肢
------	-----

選択式 (単一回答)	・ 受験者の実施利便性の向上 ・ 事務局の労力の低減 ・ 受験機会の拡大 ・ 問題の再利用が可能 ・ 感染症対策 ・ その他※
---------------	--

【Q 7】導入時の課題

回答形式	選択肢
選択式 (単一回答)	・ 導入コストが高額 ・ 問題作成数が不足 ・ 合否判定上の課題 ・ 本人確認が困難 ・ システムの煩雑化 ・ その他※

【Q 8】採用の効果

回答形式	選択肢
選択式 (単一回答)	・ 受験者の実施利便性の向上 ・ 事務局の労力の低減 ・ 受験機会の拡大 ・ 問題の再利用が可能 ・ 感染症対策 ・ その他※

【Q 9】不採用の理由

回答形式	選択肢
選択式 (単一回答)	・ 導入コストが高額 ・ 問題作成数が不足 ・ 合否判定上の課題 ・ 本人確認が困難 ・ システムの煩雑化 ・ その他※

【Q 10】提供可能な情報

記述式	
-----	--

※ 選択肢「その他」を選択した場合には、自由記述欄への記入も求めた。

(2) アンケート調査結果

イ. アンケートの回収率等

試験実施団体 67 ヶ所へアンケート調査を送付した結果、40 ヶ所（回収率 59.7%）の回答が得られた。このアンケートの概要については、表 7 のとおりである。

表 7 アンケートの概要

調査対象	国家試験等を実施している機関・団体・学校
回収件数／依頼件数	40／67 ヶ所（回収率 59.7%） （内訳：国家試験（講習なし）30／42 ヶ所、 国家試験（講習あり）4／10 ヶ所、 民間試験（講習なし）6／15 ヶ所）
調査期間	2022 年 12 月 26 日～2023 年 1 月 18 日
回答方法	① 郵送（アンケート調査票に記入し返送）：20 ヶ所 ② Web（Google フォームに回答）：20 ヶ所

ロ. アンケート調査結果

「アンケート調査の質問項目等」（表 6 参照）に対する回答結果について、Q 1～Q 2 は回答者属性（事業者名、回答者氏名、部署、連絡先など）に関する質問であるため、ここ

では省略する。Q3～Q9（Q10は回答が得られなかったため省略）の回答について、以下にて取りまとめた。

Q3～Q4 に対しての調査結果

Q3～Q4は「試験実施団体が実施している試験の実施状況」に関する質問であるため、試験実施方式（非PBT方式、併用方式、PBT方式のみ）ごとに整理し、表8にまとめた。なお、ここでいう併用方式とは、試験実施団体が実施する1つの資格試験において、非PBT方式とPBT方式の両方で実施していることを意味する。また、非PBT方式で試験を実施している団体の総数を把握するため、試験実施方式について「非PBT方式」又は「併用方式」と回答した試験実施団体の総数（10）を算出した。

表 8 資格試験の実施状況に関する調査結果

試験実施方式	非 PBT 方式		併用方式		PBT 方式	
試験を実施する 機関・団体・学校数	10 (内訳：国家試験(講習なし)4 件、民間試験(講習なし)6 件)		30 (内訳：国家試験(講習なし)26 件、国家試験(講習あり)4 件)			
試験数※23、24	7	5	50			
直近1年の受験者数	60 人～590,000 人					
本人確認の為の 事前提出物 (複数回答) ※24	公的身分証 (写真有)	3	特になし	2	受験票 (写真有)	20
	受験票 (写真有)	1	複数回答： ・ 受験票 (写真有) ・ 公的身分証 (写真無)	1	受験票 (写真無)	5
	複数回答： ・ 公的身分証 (写真有) ・ 公的身分証 (写真無)	1	複数回答： ・ 受験票 (写真無) ・ 公的身分証 (写真有)	1	複数回答： ・ 受験票 (写真無) ・ その他 (申込時に写真を提出、受験票 (写真無) は受験時)	2
	その他 (受験申込者に ID と PW を付与)	1	(無回答)	1	公的身分証 (写真有)	3
	複数回答： ・ 特になし ・ その他 (事前に提示はなし)	1			特になし	3
		複数回答： ・ 受験票 (写真有) ・ 公的身分証 (写真無) ・ 免状や資格証 (写し)			3	
		複数回答： ・ 受験票 (写真無) ・ 公的身分証 (写真有) ・ 免状や資格証 (写し)			2	
		複数回答： ・ 受験票 (写真無) ・ その他 (申込書に写真貼付有)			2	

※²³ 組織において複数の試験を実施している場合があるため、代表的な試験を最大 2 つまで回答いただいた。

※²⁴ 実技試験 1 件は除いている。

試験実施方式	非 PBT 方式		併用方式		PBT 方式	
					その他（写真票）	2
					複数回答： ・受験票（写真有） ・免状や資格証（写し）	1
					複数回答： ・受験票（写真無） ・公的身分証（写真有）	1
					複数回答： ・受験票（写真無） ・公的身分証（写真無）	1
					複数回答： ・受験票（写真有） ・公的身分証（写真有）	1
					その他（願書）	1
					その他（受験申込書（写真有））	1
					その他（受験申請書に写真貼付し、試験時に照合）	1
					その他（申込書に写真を添付させ、申込書で本人確認）	1
問題公開 （複数回答）※24	公開していない	3	公開していない	2	常に HP 等で公開している	15
	複数回答： ・公開していない ・その他（PBT 時代の問題についてのみ公開している）	2	常に HP 等で公開している	1	一定期間のみ HP 等で公開している	14
	一定期間のみ HP 等で公開している	1	過去問題集を発刊している	1	公開していない	4
	複数回答： ・公開していない ・その他（使用しなくなった問題を一部分 HP で公開している）	1	複数回答： ・常に HP 等で公開している ・その他（CBT 方式の試験は原則非公開）	1	複数回答： ・常に HP 等で公開している ・過去問題集を発刊している	3
					その他（2009 年度以降の問題を常に HP で公開している）	2
					過去問題集を発刊している	2

試験実施方式	非 PBT 方式		併用方式		PBT 方式	
					複数回答： ・一定期間のみ HP 等で公開している ・過去問題集を発刊している その他（過去 3 年分を公開している）	2 2
					複数回答： ・常に HP 等で公開している ・その他（直近 1 年分の問題を公開）	1
					複数回答： ・常に HP 等で公開している ・その他（会員専用ページで公開）	1
					複数回答： ・常に HP 等で公開している ・その他（過去 5 年分）	1
					その他（常に HP で 5 年分を公開）	1
					その他（出版社が過去問題集を発刊している）	1
					その他（試験問題の持ち帰りを可としている）	1
問題の解答形式 （複数回答）※24	選択式（択一・択多）	5	複数回答： ・選択式（択一・択多） ・記述	2	選択式（択一・択多）	29
	複数回答： ・選択式（択一・択多） ・記述	1	選択式（択一・択多）	1	複数回答： ・選択式（択一・択多） ・記述	7
	複数回答： ・選択式（択一・択多） ・実技	1	複数回答： ・選択式（択一・択多） ・記述 ・その他（論述）	1	複数回答： ・選択式（択一・択多） ・実技	3
			複数回答： ・選択式（択一・択多） ・その他（1 級は記述式）	1	複数回答： ・選択式（択一・択多） ・記述 ・実技	3
					複数回答： ・選択式（択一・択多） ・記述 ・口述	2

試験実施方式	非 PBT 方式		併用方式		PBT 方式	
					複数回答： ・選択式（択一・択多） ・口述 ・実技	2
					記述	2
					複数回答： ・記述 ・口述	1
					複数回答： ・選択式（択一・択多） ・その他（調理実技）	1
今後の試験実施 計画※24	非 PBT 方式の試験を中心に実施予定	5	PBT 方式と非 PBT 方式の試験の 両方を実施予定	5	PBT 方式の試験を中心に実施予定	37
	PBT 方式の試験を中心に実施予定	1			PBT 方式と非 PBT 方式の試験の両方を実施予定	4
	その他（CBT 方式を継続予定）	1			非 PBT 方式の試験を中心に実施予定	3
					複数回答： ・PBT 方式と非 PBT 方式の試験の両方を実施予定 ・PBT 方式の試験を中心に実施予定	2
					複数回答： ・PBT 方式の試験を中心に実施予定 ・その他（筆記試験を PBT で実施し、加えて、口 述試験と実技試験を実施）	2
					その他（CBT、IBT 方式の開発費用が捻出できない ため、当分は PBT 方式を継続する予定）	1
			その他（変更は検討していない）	1		

以上

アンケート調査の結果（表 8 参照）から、資格試験の実施状況について以下のことが分かった。

a) 試験実施方式

アンケート調査に回答した機関・団体・学校 40 件のうち、10 件（25.0%）が非 PBT 方式又は併用方式で資格試験を実施している。

資格試験単位で試験実施方式をみると、回答があった資格試験 63 件（実技試験 1 件含む）のうち、12 件（19.0%）が非 PBT 方式又は併用方式で資格試験を実施している。

資格種別で試験実施方式をみると、講習のない国家資格試験を実施している試験実施団体 30 件のうち、4 件（13.3%）が非 PBT 方式又は併用方式で試験を実施している。対して、講習のある国家資格試験ではすべての試験実施団体が PBT 方式で試験を実施しており、非 PBT 方式で実施している試験実施団体はなかった。

b) 受験者数の規模と試験実施方式

直近 1 年の受験者数の規模を見る限り、受験者数の規模と採用する資格試験の実施方式（非 PBT 方式・併用方式・PBT 方式）に明確な関係性は見られない。

c) 本人確認の為の事前提出物

非 PBT 方式及び併用方式の資格試験においては、受験票や写真付きの公的身分証の事前提出を求めていることが多いが、本人確認の為の事前提出物は「特にない」と回答している国家試験もあった。

対して、PBT 方式の資格試験でも本人確認の為に「受験票（写真有）」の事前提出を求めていることが多い。その他にも「公的身分証（写真有）」を本人確認の為に事前提出を求めている資格試験があった。

d) 問題の公開

非 PBT 方式の資格試験では、PBT 時代の問題や使用しなくなった問題のみを公開しているという回答があった。

e) 問題の解答形式

非 PBT 方式の資格試験では「選択式（択一・択多）」の解答形式が多かったが、選択式と記述式や実技を組み合わせた解答形式を採用している資格試験もあった。

Q4～Q5 に対しての調査結果

Q4～Q5 は非 PBT 方式又は併用方式の試験実施団体を対象として、非 PBT 方式の「Q4 試験システムの導入」と「Q5 試験会場の運営」に関して外部の事業者への委託の有無について質問した。回答結果の組み合わせを表 9 に整理したところ、試験システムの導入、試験会場の運営ともに外部の事業者に委託している試験実施団体が最も多かった。

表 9 試験システムの導入と試験会場の運営 (単一回答)

採用された試験システム	試験会場の運営	回答数
外部の事業者が一般向けに提供するサービスを利用	外部の事業者に運営を委託	7 ※
自組織の専用システムを構築	自組織で直接運営	1
自組織の専用システムを構築	外部の事業者に運営を委託	1
自組織の専用システムを構築	その他	2
合計		11

※現時点では PBT 方式で試験を実施しているが、令和 5 年度から併用方式で試験を実施すると回答した団体（1 件）を追加している。

また、表 9 の内容を更に精査し、外部の事業者に試験会場の運営を委託する場合には、試験システムとの組み合わせとして、表 10 のとおり 4 つのパターンがあることが分かった。

表 10 試験システムの導入と試験会場の運営の組合せパターン

組合せパターン	採用された試験システムの導入	試験会場の運営
A	二つの異なる外部業者に委託※1	
B	一つの外部業者に委託※1	
C	自組織で構築	外部業者に委託※1
D	自組織で構築	受験団体への会場運営を委託※2

※1 会場運営の外部委託には、①外部業者の契約試験会場を活用する、②各地の会場と直接契約を結ぶ、2 つのパターンがある。

※2 試験実施団体が、受験団体（組織単位で試験を受験する団体）に試験の会場運営を直接委託すること。

Q6及びQ8に対する調査結果

Q6及びQ8は、非PBT方式の試験を「Q6採用の理由」と「Q8採用の効果」に関する質問については、同一の選択肢及び回答形式（選択式〔単一回答〕）を設定した。それぞれの質問への回答結果を表11と表12に整理して比較したところ、採用の理由と採用の効果の両方で、受験者にとってのメリットである「受験者の実施利便性の向上」、「受験機会の拡大」を選択している試験実施団体が多かった。

表 11 非 PBT 方式を採用した理由（単一回答）

非 PBT 方式の採用理由	回答数
受験者の実施利便性の向上	5
受験機会の拡大	4
問題の再利用が可能	1
感染症対策	0
その他	0
合計	10

表 12 非 PBT 方式を採用した効果（単一回答）

非 PBT 方式の採用効果	回答数
受験者の実施利便性の向上	6
受験機会の拡大	2
問題の再利用が可能	1
その他	1
感染症対策	0
合計	10

「その他」の回答として以下の記載があった。
・初めから CBT 方式を採用していた。

Q7及びQ9に対しての調査結果

Q7及びQ9は、非PBT方式の試験を「Q9採用していない理由（不採用理由）」と「Q7導入時の課題」に関する質問については、同一の選択肢及び回答形式（選択式〔単一回答〕）を設定した。前者はPBT方式のみを実施している試験実施団体を対象とし、後者は非PBT方式を導入している試験実施団体を対象とした質問であり、それぞれの回答結果を表13と表14に整理したところ、以下のことがわかった。

- ・ 非PBT方式の不採用理由と導入時の課題として「導入コストが高額」と回答した試験実施団体が多かった。
- ・ 非PBT方式の不採用理由として「本人確認が困難」と回答した試験実施団体は多かったが、非PBT方式の導入時の課題として回答した試験実施団体はなかった。

表 13 非PBT方式の不採用理由（単一回答）

非PBT方式の不採用理由	回答数
導入コストが高額	7
本人確認が困難	7
その他	5
合否判定上の課題	3
問題作成数が不足	2
システムの煩雑化	2
（無回答）	3
（無効回答）	1
合計	30

「その他」の回答として以下の記載があった。

- ・ 関係省庁の方針による。
- ・ 製図試験を非PBT方式で行うことが困難。
- ・ 非PBT方式を必要としないため。
- ・ すべて該当（導入コストが高額、問題作成数が不足、合否判定上の課題、本人確認が困難、システムの煩雑化）

表 14 非PBT方式の導入時の課題（単一回答）

非PBT方式の導入時の課題	回答数
導入コストが高額	4
問題作成数が不足	3
その他	2
合否判定上の課題	1
本人確認が困難	0
システムの煩雑化	0
（無回答）	0
（無効回答）	0
合計	10

「その他」の回答として以下の記載があった。

- ・ 運用の安定感（導入初期の為）
- ・ 問題の量産

3. 2 ヒアリング調査（試験実施団体）

（1） ヒアリング調査の概要

アンケート調査の結果を基に、試験実施団体に対して、更に詳細な情報を収集するためのヒアリング調査を行った。ヒアリング調査の概要については、表 15 のとおりである。なお、ヒアリング対象については、異なる国家試験並びに技術検定及び検定に類する試験の取組みを調査できるよう配慮した。

表 15 試験実施団体に対するヒアリング調査の概要

調査対象	国家試験等を実施している機関・団体・学校
調査先数	10 ヶ所 (内訳：非 PBT 方式（国家試験）3 ヶ所、非 PBT 方式（民間試験）3 ヶ所、PBT 方式（国家試験）4 ヶ所)
調査期間	2023 年 1 月 31 日～2023 年 2 月 22 日
調査方法	① 対面 ② Web 会議システム
調査時間	約 1 時間／1 ヶ所
調査項目 【対象】	イ. 非 PBT 化の理由・背景【非 PBT 方式実施団体のみ】 ロ. 採用した試験システム【非 PBT 方式実施団体のみ】 ハ. 非 PBT 化しない理由【PBT 方式実施団体のみ】 ニ. 試験運営の方針（本人確認、不正行為の防止、作問・採点方式）【両方】 ホ. PBT 化への移行プロセス【非 PBT 方式実施団体のみ】 ヘ. その他【両方】

（2） ヒアリング調査の結果

試験実施団体ごとに行ったヒアリングは以下のとおりである。

イ. 非 PBT 化の理由・背景に対する回答

試験実施団体が非 PBT 化に取り組んだ理由及び背景として、以下のとおりコロナ禍の影響が大きかったことがわかる。

- コロナ禍により、PBT 試験の実施が中止や延期に追い込まれ、事業継続の必要性に迫られて CBT/IBT 化した。
- コロナ禍による下記 3 つの要因も CBT/IBT 化に大きく影響している。
 - ① 学校の閉鎖などによる使用可能な会場数の減少によって、試験会場を探す労力の増加
 - ② ソーシャルディスタンス（三密回避）による使用会場数の増加に伴う費用の増加
 - ③ 以下の理由による人件費の増加
 - ・人との接触を避けたいという状況下での試験監督官のなり手不足
 - ・使用会場数の増加による試験監督官の人数増

ロ. 採用した試験システム

試験実施団体の採用した試験システムの特徴及びその採用理由について、以下に整理した。

- 各団体が採用した試験システムについては、大きく以下2つのパターンに分けられる。
 - 1) 自団体の仕様に合わせ、システム開発業者へ発注し、独自のシステムを構築した。
 - 2) 非 PBT 事業者のクラウドシステムを採用した。
 - ① 申込システムから合否通知まで、試験の全てフローで非 PBT 事業者のクラウドシステムを使用している。
 - ② 試験配信システムについてのみ非 PBT 事業者のクラウドシステムを使用し、それ以外の申込・合否通知等については自団体の独自のシステムを使用している。
- システムの採用の経緯としては、仕様を固めて入札等で決定したという試験実施団体は少なく、コロナ禍における緊急対応であるため時間的制約が大きく、既取引のあった PBT 事業者の他部署または関連企業等に発注していることが多い。
- 仕様を固めてから数社に対して試験システム提案を求めて採用するシステムを決定した試験実施団体は、非 PBT 事業者の採用理由として以下を挙げた。
 - ・実績
 - ・レスポンスの良さ
 - ・コスト
 - ・運用開始までのスピード等

ハ. 運営の方針

試験実施団体における試験運営の方針については、「a) 本人確認」「b) 不正行為の防止」「c) 作問・採点方式」の観点から以下のとおり整理した。

a) 本人確認

本人確認の方法については、「試験申込時」「試験当日」に分けて具体的な取組例を以下に記載する。

<試験申込時（事前）>

- 郵送による申込みの場合：紙の願書(申込書)に決められたサイズの顔写真を添付する。
- インターネット申込みの場合：顔写真を申込システムにアップロードする。

国家資格試験（講習ありの場合）の一例：

申込時の本人確認の方法は、

- ① 受験者が、申込みシステム上に顔写真データをアップロードする。
- ② 試験実施団体が、法令で受験に必要と定められた講習の受講履歴に基づく受験資格の有無の審査と、アップロードされた当該講習の修了証を組み合わせ確認する。

<試験当日>

- PBT 試験の場合：

受験者は紙の受験票に顔写真を貼り、試験当日持参する。試験会場において当該受験票を用いた本人確認と受験票の回収を行う（合格の場合は、受験票で合格証（免許等）発行に伴う本人確認も行う）。

● CBT 試験の場合：

本人確認の方法は、CBT 事業者が定めた試験会場の運営基準に沿って実施している団体が多い。本人確認方法の具体例は、以下のとおり。

- ・ 顔写真付きの公的身分証明書(自動車運転免許証、パスポート、マイナンバーカード等)の提示による確認。
- ・ 受験者のうち学生の占める割合が高い資格については、学生証や保険証等を本人確認書類として活用して確認。
- ・ 試験申込時に付与した ID とパスワードのみで確認。

● IBT 試験の場合：

Web カメラで試験開始前に本人の顔写真を撮影し、試験中にも Web カメラ越しに定期的に本人確認を行う。

b) 不正行為の防止

試験実施団体における不正行為防止に関する考え方や取組み状況について、以下に記載する。

- CBT 試験における不正行為を防止するために、非 PBT 事業者が定めた基準に沿って運営している団体が多い。但し、各団体において、試験会場の事前下見や、問題・苦情が発生した場合にテストセンターに対して事業者がどのような指導を行っているのかを抜き打ちで確認する等の対応は行っている。
- IBT 方式を採用した試験実施団体においては、試験監視のための独自の基準を採用している場合がある。(例：画面 [Web 検索やコピー、スクリーンショット等の操作)の有人監視を非 PBT 事業者選定の必須条件にする等])
- IBT 方式の試験については、不正行為の防止が難しいと考えている団体もある。

c) 作問・採点方式

試験実施団体における作問や採点に関する考え方や取組み状況について、以下に記載する。

- 作問については、法令等に基づき、条件を満たした方や有資格者、監督官庁の指名者・推薦者からなる委員会で行っている。
- 作問の際に問題の揺らぎや他の問題のヒントになるような問題の排除等を行っている団体もある。
- 簡単な統計データを用いて試験結果を分析している団体はあるが、IRT(項目反応理論)に対応している団体は少ない。
- 作問の精査の際に、非 PBT 事業者から準備すべき問題数や問題の揺らぎの排除等のコンサルティングを受けている団体もある。
- PBT と非 PBT を併用している試験に関しては、同一実施回において同じ問題を使用せず、PBT の問題を翌年改編して非 PBT の問題プールに追加している団体もある。

- 試験を実施した後、委員会において、極端に難しい問題やひっかけ問題がないかを確認し、受験者の解答を照らし合わせ、可否をオーソライズしている団体もある。

二. 非 PBT 化への移行プロセス

試験実施団体における非 PBT 化への移行プロセスについて、移行パターンと苦勞した点を以下に整理した。

- 今回ヒアリングした試験実施団体における非 PBT 化への移行のパターンは、各団体が試験運営の方針や試験の有用性をどの様に考えるか等によって異なり、申込方法と試験実施形式の観点から下表のとおり 4 種類に整理できる。

非 PBT 化への移行パターン

移行 パターン	申込方法		試験実施方式	
	インターネット	郵送	非 PBT	PBT
1	○	○	○CBT/IBT 併用	○
2	○	○	○CBT のみ	○
3	○	×	○CBT/IBT 併用	×
4	○	×	○CBT のみ	×

○：採用 ×：不採用

- 移行プロセスにおいて苦勞した点として、主に以下が挙げられた。
 - ① 作問(問題数と公平性の確保)
 - ② 既存の自社独自の申込システムと非 PBT 事業者の試験配信システムの連携等
 - ③ インターネット申込みに対する受験者からの問い合わせ対応

ホ. 非 PBT 化しない理由（ヒアリング対象：PBT 試験実施団体のみ）

PBT 方式の試験のみを実施している団体が非 PBT 化へ移行していない理由について、以下に整理した。

- CBT 導入時のコスト（既存の申込システム等とのシステム連携等にかかる初年度費用が大きな障壁となっている）
- 複数の資格と多数の受験科目の組み合わせを管理するシステム（自社システム）と試験配信システム（外部業者のシステム）を連携すると、業務の煩雑化が予測されるため CBT 化のメリットが見込めない。
- 非 PBT 試験において不正行為(カンニング等)が防止できるのか不安がある。
- コストと労力の両面で、試験を複数回実施するために必要な問題数を用意するのが難しい。
- CBT 試験を年 1 回のみ実施する場合、会場や機材（PC 等）の用意に莫大な手間と費用

が掛かるため、非 PBT 化する場合は IBT が前提になる。しかし、IBT は不正行為の防止に難があると考ええる。

- 年間に試験を複数回実施した場合、試験の公平性を担保できるか不安である。
- CBT の導入を検討したことがないため、コスト、作成すべき問題数、合否判定基準、本人確認方法、システムの煩雑化のうち、何が実際に導入障壁となるのかわからない。
- 二次試験は論文問題のため採点のシステム化が困難であることから、二次試験の CBT 化は全く検討していない。中途半端に一次試験のみを CBT 化すると、かえってコストがかかると考えている。

へ. その他

その他、試験実施団体から得られた回答について以下に記載する。

- 天候不順等の災害で試験が中止になった場合は、PBT 試験と比較して代替日程の確保が容易である非 PBT 試験の方が、受験者にとって利便性が高い。
- PBT と CBT を併用している試験の場合は、CBT 試験に対する受験者の誤解がある。(PBT 試験の方が「格」が高いと思っている方がいる)
- 受験者都合による非 PBT 試験の欠席については、返金していない。

3. 3 非 PBT 事業者に関する整理・分析

(1) 非 PBT 事業者へのヒアリング調査の概要

資料調査（一次調査）の結果を踏まえ、国内の非 PBT 方式の試験システムや試験運営サービスについて更に詳細な情報を収集するために、国内の非 PBT 事業者 4 社に対してヒアリング調査を行った。ヒアリング調査の概要については、表 16 のとおりである。

表 16 非 PBT 事業者に対するヒアリング調査の概要

調査対象	試験システムや試験運営サービスを提供する非 PBT 事業者
調査先数	4 ケ所
調査期間	2023 年 1 月 27 日～2023 年 2 月 9 日
調査方法	① 対面 ② Web 会議システム
調査時間	約 1 時間／1 ケ所
調査項目	イ. 提供されている試験システムの概要 ロ. 提供されている試験会場運営サービスの概要 ハ. 試験の公平性を担保するための仕組み ニ. 申込方法 ホ. データの管理 ヘ. アクセシビリティへの配慮 ト. 非 PBT 化への移行対応 チ. 非 PBT 化にかかるコスト ニ. その他

(2) ヒアリング調査の結果

今回のヒアリング調査結果から、殆ど全ての非 PBT 事業者は、試験システムと試験運営サービスの両方を提供していることがわかった。そこで、上記のヒアリング項目に準じてヒアリング調査の結果とインターネット上の公開情報等を基に、以下に整理した。

イ. 提供されている試験システムの概要

各事業者が提供する試験システムの特徴が分かれるポイントについて、以下のとおり「事業者の方針」「システムの機能」の観点から整理した。

【事業者の方針】

試験システムに関する事業者の方針には、大きく分けて「厳格性重視」と「利便性重視」があった。

- 厳格性重視：国家試験のようなハイ・ステークスな試験にも活用できるよう、厳格な試験運用、試験コンテンツの保護・管理等を重視している。

- 利便性重視：講習後の習熟度確認、組織内試験等での利用を目的とし、厳格さよりは、受験者と試験実施団体の利便性や手軽さ(価格・操作性)を重視している。
また、試験システムには、大きく分けて「ワンストップ型」と「機能特化型」があった。
- ワンストップ型：申込～受験料の支払い～試験配信～結果提示のシステムを全て1つのパッケージとして提供可能。
- 機能特化型：試験配信に特化したシステム提供で、他の様々なシステム(申込、決済等)との連携可能。

【システムの機能】

試験システムの機能の特徴が分かれるポイントは、以下のとおりである。

- 出題形式の多様さ(多肢選択式問題、記述式問題、画像や計算式の表示、パフォーマンス測定が可能なリスニング・スピーキング等)
- 試験作成の簡便さ(試験実施団体によるシステムへの問題の搭載が可能かどうか等)
- 他のシステム(申込システムや講習システム(LMS)等)との連携
- 試験システムに同時アクセスできる受験者数の違い
- 本人確認や試験監督の厳格性・仕組み

ロ. 提供されている試験運営サービスの概要

各社が提供する主な試験サービスは、以下のとおりである。

- 問題作成のアドバイス
- 問題の試験配信システムへの登録
- 受験票の発行
- 申込問合せ対応
- 受験費用徴収
- 試験会場提供(CBT事業者が確保している会場数は全国に50～130か所前後)＊1
[PC等の機材・監視のためのハード(カメラ、間仕切り等)及びソフト(マニュアル、試験官訓練)を含む]
- 試験の受付・監督業務
- 不正行為防止のための対応(受験者の画像撮影、カメラ貸し出し等)＊2
- 受験時のトラブル対応(通信障害等)
- 採点・集計業務
- 複数の試験版を用いる場合、版を跨いだ難易度等の標準化等を支援
[出題した試験問題が機能しているかを分析し、問題の再利用の可否についての評価を含む]
- 合格証の発行、合格カード作成
- 受験者データの管理・バックアップデータの保管

＊1：CBT試験の場合のみ

* 2 : IBT 試験の場合のみ

ハ. 試験の公平性を担保する仕組み

試験の公平性を担保するために、各社が行っている取組みの具体例は以下のとおりである。

【試験当日の本人確認】

- 本人確認書類は2点：公的身分証（写真付き）1点＋社員証・学生証（写真なしでもよい）1点
- 手のひら静脈スキャンや顔写真撮影を含むデジタル ID 認証
- 予約者情報と公的身分証の情報との照合
- 公的身分証（写真付き）1点で本人確認
- 当日、顔写真（デジタル）を撮影し、会場の受付端末へのデジタル署名をとる。

【試験監督】

<CBT/IBT 試験共通>

- サブ画面への移動を検出し、テストを一時停止する機能がある。
- コピー＆ペースト、マウス操作(右クリック)、キーボードショートカット及びプリントスクリーン（印刷）を防止している。

<CBT 試験>

- 不正行為と判断した場合、試験監督が受験者の試験システムをシャットダウンできる機能がある。
- 受験者コントロールとして、各問題・テストの解答時間の設定及び解答後の変更の禁止/許可の設定ができる。
- 試験時に使用したメモ用紙は持ち出せない規則にしている（問題漏洩防止のため）。
- 会場の監視カメラは2席に1台設置している。

<IBT 試験>

- Web カメラで試験前に本人写真を撮影し、試験中は人によって定期的に本人確認を行っている。
- 30 秒に1枚受験者の画像を撮影(30 秒以内ではカンニング行為はできないという想定)。1 試験1名 360 枚ほどの画像を撮影し、試験終了後に不正行為が行われていないかそれらの画像を目視で確認している(360 度カメラの貸出し)。
- 受験中に試験以外の画面を表示できないよう、試験画面を固定している。

二. 申込方法

各社とも試験実施団体のニーズに合わせ、以下とおり様々な申込に対応している。

- 個人でも、団体でも申込可能。
- 試験実施団体の申込システムと各社が提供する試験配信システムの連携が可能。

- 受験料の支払いは、様々な決済方法に対応可能。
 - ・コンビニ払い ・銀行振込 ・クレジットカード ・Pay-easy

ホ. データの管理

事業者によって受講者データの管理方法等かなり違いがあった。以下に主な具体例を挙げる。

- 受験者の受験履歴だけでなく、講習履歴のデータ管理も行うことができる。
- クラウド上の安全を確保した環境において永続的に結果データを保存し、削除はしない。試験実施団体はそこからダウンロードする形でデータをいつでも活用できる。
- 受験者情報の管理については試験実施団体が保有する学習管理システム（LMS）と連携し、LMS 側で定義されたユーザー情報との相互認証という形をとるため、CBT 事業者側では仮名の状態で受験できる。受験した情報を LMS に戻すことで個人が特定されるため、CBT 事業者内のシステムに個人情報を持つ必要がない。

ヘ. アクセシビリティへの配慮

アクセシビリティについては対応が様々であり、試験システムが WCAG²⁵（Web Content Accessibility Guidelines）に対応している事業者から、まだ試験会場におけるバリアフリーや PC の操作補助者の同伴の許可等、身体的な障害にのみにしか対応できていない事業者まであった。

しかし、全ての事業者において、アクセシビリティ向上の必要性は十分認識し、受験機会の公平性拡大に対する努力は行っている。具体的な取組みは以下のとおりである。

- 2023 年 4 月から、試験予約のサイトの静的ページについては、Web アクセシビリティ配慮したデザインに変更する予定である。
- 合理的配慮として、杖や補聴器具の持込み等にも対応する。
- 弱視の方には表示内容を拡大する機能、色盲の方には見やすい色に変更等の機能を試験配信システムに装備している。
- IBT という試験実施方法は、ある意味身体的障害の方には利便性が高いと認識している。
- 受験者から事前申告があれば、可能な限り対応する。

ト. 非 PBT 化への移行対応

殆どの非 PBT 事業者が、全体的な試験システム構築の観点からの助言・支援が可能であ

²⁵ WCAG は、Web コンテンツをよりアクセス可能にするための幅広い推奨事項を定めたガイドラインである。これらの推奨事項に基づいた Web コンテンツを制作することにより、失明や弱視、難聴、運動制限、言語障害、光線過敏症、学習障害等、より幅広い障害を持つ人々が Web コンテンツにアクセスしやすくなる。但し、これらの障害を持つユーザーの全てのニーズに対応するわけではない。

参考：Web Content Accessibility Guidelines (WCAG) 2.1 (<https://www.w3.org/TR/WCAG21/>) 一部翻訳 (2023/2/14 時点)

る。3. 2 (2) ニ (非 PBT 化への移行プロセス) でも記載したが、非 PBT への移行の中で、問題数の確保と公平性の担保の 2 つが試験実施団体にとっての大きな負担であることがわかっている。非 PBT 事業者は、作問に関するこれら 2 つの負担のうち公平性の担保という観点からの試験実施団体への支援に注力している。公平性の担保に対する支援の具体例を以下に挙げる。

- 日本に 1 名サイコメトリシャン (教育心理学の専門家) が在籍し、試験問題の公平性の検証ができる。
- 試験実施団体と一緒に作問する専門家 (テストデベロッパー) を擁している。
- 試験実施団体が作成した問題をシステム上に載せて CBT 化する際に、IRT 等のテスト理論に基づいたテスト設計が必要であれば、教育測定学や教育心理学等の専門家の協力を得ることは可能である。
- 作問者ごとに問題の難易度等のバラつきがないよう、作問時に押さえるべきポイント等について提示し、修正作業の少ないプロセスになるようサポートする。
- 問題の質 (困難度、識別力等) を評価する。
- 過去の PBT 試験の問題を再利用するために統計処理し、公平性の検証を検証する。
※ 1 問あたり 300 個のデータが必要である。IRT を行う場合には更に大量のデータが必要である。

チ. 非 PBT 化にかかるコスト

非 PBT 化にかかるコストについては、金額だけでなく、労力や有益性も考えると多様なケースが想定されるという事で、具体的な数字は確認できなかった。以下はコストに関する面談者の言葉をそのままに近い形で記載したものである。

【金銭面】

- 金銭的な面から考えると、受験者が年間 10 万人レベルの試験であれば、PBT より CBT の方がコストは抑えられる。
- PBT の場合、試験用紙の印刷、輸送、事務局の人件費、会場手配、移動費等が発生する。
- 受験者が 100 名に満たない試験の場合は、投資対効果という観点から非 PBT 化は割に合わない可能性が大きい。
- 非 PBT 化について多様な契約プランを用意しており、PBT と非 PBT のどちらがコストパフォーマンスがよいのか断定はできない。

【労力面】

- 受験者の労力 (会場選択、日時選択等) については、CBT 化によって低減され则认为られる。
- ある程度の試験規模であれば、試験実施に従事するスタッフの移動時間、一斉受験の手間を考えると IBT の方がトータルコストは下がると思う。

【有益性】

- 非 PBT 化することで様々な分析が可能になるため、試験の価値・質の向上を図ることができる。非 PBT 化は金銭では測れない有益性がある。

二. その他

その他、非 PBT 事業者へのヒアリングのなかで得られた「出題形式」「問題漏洩」に関する貴重な見解を以下に記載する。

【出題形式について】

代表的な出題形式には以下 2 つがある。

- 1) プール形式 プールした問題から適時出題する。(プールすべき問題数は、アカデミックな試験は出題数の 12 倍、資格試験は 7 倍必要と言われているが、凡そ出題数の 10 倍は必要であると言われている。)
- 2) フォーム形式 毎年テストセットを複数用意し、次回受験可能となるまでの間隔(例: 6 か月等)を指定して同一問題の再出題を防ぐようにする。(通常 3 セット用意すると良いと言われている)

<実施例>

- 受験可能なのは、前回の受験から 6 カ月後以降とする。
- 1 回目の試験でテストセット A が使用されたら、次回の試験ではテストセット B が使用される。
- 毎年テストセットを 3 つ作ることににより、同じ問題を再出題することはない。

【問題漏洩について】

- 問題漏洩が発生する可能性は、問題数や覚えにくさなどの要素によっても左右される。

(3) PBT 方式と CBT 方式の試験に係る費用比較シミュレーション

チ. で述べたとおり、非 PBT 化にかかる費用のシミュレーションは変数(試験実施の前提: 受験料、受験者数、受験回数等)が多すぎるため、数字で提示することは困難である。そこで、本項においては、PBT 方式と CBT 方式の試験にかかる費用項目を比較することとした。

比較に先立って、まずは試験実施にかかる費用項目を整理するために、「試験システム」と「試験運営サービス」を以下のとおり定義した。

- 試験システム: 試験実施、申込受付、受験料の決済、試験結果の採点・通知に関わるシステムの設計・開発・運用
- 試験運営サービス: 試験の実施に関して、受験者と接するサービス及びそれに付随する業務(会場手配、申込受付、問合わせ対応、試験監督等)

加えて、試験実施の方法等によって試験にかかる費用項目(工数)も変化するため、費用項目を比較するための条件として以下5つを設定した。

条件①：CBT 方式は、試験システム開発会社の既存システムを利用して実施。(特殊なカスタマイズは含まない)

条件②：問題は全て多肢選択形式。(記述式は含まない)

条件③：CBT 方式のみ試験運営サービス(申込受付、試験実施、[会場手配・試験監督])を利用。PBT 方式は試験運営に関する全ての事項を試験実施団体が自ら手配。

条件④：PBT 方式は問題を再利用しない。CBT 方式は問題を再利用かつ新たな問題を追加。

条件⑤：CBT 方式はテスト理論に基づくテスト設計を実施。

上記の定義及び条件に基づき、表 17において、試験を PBT 方式及び CBT 方式で実施する場合の費用項目を整理し、試験実施団体が業務を担う項目には「●」印、非 PBT 事業者が委託を受け業務を担う項目には「○」印、業務として発生しない項目には「—」印を付した。

表 17 PBT 方式・CBT 方式の費用項目比較表

区分	費用項目	PBT 方式	CBT 方式
試験システム	システム関連		
	申込受付システムの初期導入	—	○
	申込受付システムの維持・管理	—	○
	決済システムの初期導入	—	○
	決済システムの維持・管理	—	○
	テストシステムの初期導入	—	○
	テストシステムの維持・管理	—	○
	受験者データの管理	●	○
	試験の編集		
	問題データ化(システムへの登録作業)	—	○
	試験の質の担保のためのコンサルテーション(試験結果の分析、テスト理論に基づいた助言、試験の質向上のための提案等)	—	○
試験運営サービス	申込受付		
	受験料の納付確認	●	○
	受験票発行	●	○
	受験票送付(人件費・切手費用)	●	○
	受験者登録作業(紙媒体からの入力)	●	—
	問い合わせ業務	●	○
	試験実施		
	テスト・解答用紙の印刷	●	—
	テスト・解答用紙の配送	●	—

	会場の手配（人件費・会場費用）	●	○
	会場の設営（人件費）	●	○
	什器・備品の手配（デスク等）	●	○
	受付・本人確認（人件費）	●	○
	試験監督（人件費）	●	○
	試験実施補助員（問題配布）（人件費）	●	—
	採点・結果通知		
	結果集計・採点業務	●	○
	合格証作成	●	○
	合格証発送（人件費・切手費用）	●	○
試験問題 について （試験実施 団体必須）	試験問題の作成・整理		
	作問	● 作問数少	● 作問数多
	出題された問題の整理	● 整理数少	● 整理数多

【凡例】 ●：試験実施団体の役割（人件費・経費）

○：非 PBT 事業者の役割（システム運用含む）

—：業務として発生しない

4 非 PBT 方式による国家試験等の在り方の検討

4. 1 非 PBT 方式で国家試験等を実施する場合の実施内容

現在 PBT 方式で実施している試験を非 PBT 方式に移行する際に、試験実施団体にとって最善の選択が何であるかについては、各団体や試験の置かれている状況や特性を多面的に検討し、判断する必要がある。本項では、非 PBT 方式への移行について、試験の厳格性と利便性、蓄積すべき問題数、受験者の属性の観点等から整理した。

なお、本報告書において、試験の実施方式は、1.5 (1) で述べたとおり、大きく分けて PBT 方式と非 PBT 方式の 2 つに整理し、非 PBT 方式は更に CBT 方式及び IBT 方式に分類している。試験の実施方式については、表 1 に詳細を定義しているので、以下に一部再掲する。

表 1 8 試験実施方式の定義（一部再掲）

実施方式		特徴
PBT 方式		方法：紙面（問題・回答用紙等）を配布・回収する方法 場所：指定された会場 時間：全国一斉に実施
非 PBT 方式	CBT 方式	方法：コンピュータ端末（会場に設置のもの）を使用する方法 場所：指定された会場（テストセンター） 時間：①個別の時間帯で実施 又は ②全国一斉に実施
	IBT 方式	方法：コンピュータ端末等（任意のもの）を使用する方法 場所：任意の場所 時間：①個別の時間帯で実施 又は ②全国一斉に実施

(1) 試験の実施方式

国家試験等の非 PBT 化の検討という要請に対して、最も大きな方向性となるのが、試験の実施方式である。国家試験のような厳格性が求められるハイ・ステークスな試験においては、替え玉受験、カンニング、問題漏洩等の不正行為を防止するために、受験環境（会場、使用端末の設定等）の統制や試験セキュリティ（本人確認、試験監督）の確保が重要課題である。

また、本事業で実施したアンケート調査において、試験実施団体が、非 PBT 方式の試験を採用した理由の一つとして挙げたのが「受験機会の拡大」及び「受験者の利便性向上」であった（3. 1 (1) Q6、Q8 の回答）。

そのため、ここでは「イ. 受験環境の統制及び試験セキュリティの確保」と「ロ. 受験者の受験機会拡大及び利便性向上」にポイントを置き整理した。

イ. 受験環境の統制及び試験セキュリティの確保

CBT 方式の試験では、会場において PBT 方式同様に身分証明書（顔写真付）による厳格な本人確認や入退室管理を行うことができるため、替え玉受験の可能性は、IBT 方式と比べて低いといえる。加えて、非 PBT 事業者によって使用端末の機能設定や会場での有人試験監

督が行われているため、カンニングや問題漏洩の可能性も、IBT 方式と比較して低いといえる。

一方、IBT 方式においては、受験者に対して、不正行為を防止するために受験環境及び使用端末の準備に関する様々な要件を満たすことを義務として課し、あわせて試験セキュリティを高めるための本人確認の仕組みや監視システムを開発、導入している。しかし、AI システムや Web カメラ越しでの確認、監視となるため、現時点においては技術的な不確実性が払しょくできない。

また、令和 2 年度及び令和 3 年度の報告書によれば、非 PBT 方式の試験の実施方式として、CBT 方式及び IBT 方式が検討及び不正行為の監督がより困難であると考えられる IBT についての実証を行っている。その結果、不正行為の監督に関する技術は未だ発展途上であり、人の目による監督の必要性が認められるとともに、カメラを通しての監督ではカメラの死角が不正行為を容易にする可能性があることについても言及された。

ロ．受験者の受験機会拡大及び利便性向上

PBT 方式の試験は、その実施に多くの人的な工数と経費がかかるため、試験の実施回数と会場に限られる場合が多い。しかし、非 PBT 方式では、通年や 1 か月等の期間を限定して試験の実施回数を増やすことができる。また、非 PBT 方式のうち、CBT 方式では受験者は試験会場の選択肢のなかから自身に便利な会場で受験することができ、IBT 方式では受験者が受験環境を自由に選択できるため自宅や職場等で手軽に受験することが可能である。非 PBT 方式の試験は、PBT 方式よりも受験者の利便性が高く、なかでも IBT 方式は CBT 方式よりも受験の日時や環境に関する選択肢の幅が広く、自由度が高い。

上記より、試験の厳格性と受験者にとっての利便性はトレードオフ関係にあり、試験の不正行為防止等の厳格性が求められるということは、一方で受験環境の一定の制限を行う等の利便性が損なわれるということである。

したがって、ハイ・ステークスな試験を非 PBT 化するのであれば、受験者の利便性よりも試験の厳格性を優先する必要があるため、CBT 方式での試験実施の方が望ましい。

(2) 出題方式

高圧ガス保安法における国家試験等では、各受験者に同じ問題を出題することで、出題問題に由来する不公平が起こらないように運用している。すなわち、各試験で使用される問題セットは 1 種類のみということである。この運用は古典的試験理論において、異なる試験間の結果比較が困難であるという課題に対しての一つの解決策として機能しているが、受験者全員に同じ問題を出題するためには様々な制約がある。その一つが、問題を事前に特定されてはいけないというものであり、これを守るために試験の実施日時を揃え、問題の管理に注力した運用をしている。また、試験を運営、監督するために同一日時に稼働させることができる人的リソースは限られているため、効率的に運営するために大規模な試験会場を設定するなど、試験運用の様々な側面に影響を与えている。

CBT 方式の利点を活かした試験の運用を考える場合、PBT 方式では実施することが難しい複数の問題セットを用いた試験の実施方法がある。古典的試験理論に基づく同一問題の出題における

欠点を解決する仕組みを利用することが可能であり、コンピュータにより問題をランダムに出題することとした場合、出題される問題は試験直前まで特定することができず問題流出によるリスクは大幅に低下する。また、試験問題の印刷、配送、採点作業などの事務作業が大幅に削減できる可能性が高く、この事務作業の効率化のために必須であった大規模会場の確保も不要になる。

したがって、高圧ガス保安法における国家試験等の CBT 方式への移行に際しては、問題データベースからのランダム出題を念頭に検討することが望ましい。

（３）問題形式

高圧ガス保安法における国家試験等では、主に選択式問題と記述式問題の２種類を取り扱っており、記述式問題は各免状種類の最高難易度の試験²⁶において出題をしている。そこで以下では「イ．選択式問題」と「ロ．記述式問題」に関してそれぞれ検討した。

イ．選択式問題

選択式問題は記述式問題の欠点を解決するため、また大規模試験を実施するために客観式問題が利用されるようになり、コンピュータ処理により採点者の属人性を排除した状態で効率的に結果を求められる点で優れている。

高圧ガス保安法の試験においても、大部分がこの形式で出題されているが、同法では、１つの問題を４つの設問で構成しており、設問内容の正誤の組み合わせを解答させる形式となっている。この形式の問題点としては、４つの設問を解かせる正誤の組み合わせは 16 パターン考えられるが、実際の問題では５つの組み合わせパターンのみを選択肢として与えており、この段階で 11 通り（79%）の解答パターン（自由度）を排除している。

一方、IRT を活用した試験では出題される問題は、IRT による事前評価があることを前提としているところ、問題の運用中に問題を構成する４つ設問の１つに変更の必要性が生じた場合、他の３つの設問を含めたこの問題の事前評価は失効する。すなわち、１つの設問に対する変更が他の３つの設問の運用にも影響を及ぼすということになる。これは、IRT のような統計データの積み上げが精度に反映される形式の運用において、データの損失が大きすぎ（400%）、運用に支障をきたすと考えられる。また、同じ出題分野の中で、設問構成の入れ替えなどもできず、運用の自由度に欠ける。

したがって、これまで４つの設問から１問を構成していた選択肢問題に関して、これを解体し、設問単位での出題とする方向で検討をしたい。

（例題）高圧ガス保安法における選択式問題の出題形式のイメージ

Ⅰ．食品の名称に関する次の記述のうち正しいものはどれか。

イ．エシャロットとエシャレットは名称は異なるが同じ野菜である。

ロ．丸十とは、薩摩藩島津氏の家紋にちなんだ野菜の名称であるが、別名を薩摩芋という。

²⁶ 甲種化学製造保安責任者免状、甲種機械製造保安責任者免状、第一種冷凍機械製造保安責任者免状の３つ。これらの免状と下位の免状の差は、主に事業所で取り扱うガス数量である。

- ハ、ジン・トニックのジンはジンジャーのことであるため、ジンジャーハイボールがメニューにない場合には、ジン・トニックを注文することで代替できる。
- ニ、ジン・トニックはアルコール度数が高いため、水割りで注文すると良い。

- (1) ロ (2) イ、ハ (3) ロ、ニ
(4) イ、ロ、ニ (5) ロ、ハ、ニ



1. エシャロットとエシャレットは名称は異なるが同じ野菜である。
2. 丸十とは、薩摩藩島津氏の家紋にちなんだ野菜の名称であるが、別名を薩摩芋という。
3. ジン・トニックのジンはジンジャーのことであるため、ジンジャーハイボールがメニューにない場合には、ジン・トニックを注文することで代替できる。
4. ジン・トニックはアルコール度数が高いため、水割りで注文すると良い。

ロ. 記述式問題

記述式問題は、口頭での質疑に次いで古い問題形式であり、大きく分けると小論文式と短答式に分けられる²⁷。小論文式では、問題に対する正解はなく、修辞、論理、内容の適切さ等を評価することが多く、解答までのガイドラインを与えない点で、受験者が自身の力で正答を導けるかどうか、表現力や思考の自由度を与えており、また多くの場合必要知識を暗記しているかどうかについても測定することができる点で優れた問題形式である。

短答式では、望まれる正解が設定されていることが多く、この場合は表現力や思考の自由度より、必要知識の暗記、思考力（網羅性、必要十分条件等の検証力）に重きが置かれていると考えられる。これら記述式問題では一般に解答時間を長めに設定する必要がある、選択式問題と比較して問題数を限定することから、試験範囲を十分に網羅した試験にならないことが多い。また、受験者の得点における偶然誤差に関しても、「採点結果が採点者間で一致しない（採点者による誤差）」、「採点結果が、採点者自身においても一致しない（採点者自身によるブレ）」という問題点が古くから指摘されており、対策として「同一の問題は同一の採点者が採点する」、「判断が難しいものは協議をする」等が実施されているが、この偶然誤差は試験の妥当性、信頼性を下げる要因であることから、この点で客観式の問題へ変更する等の対策がとられ、今日までに数を減らしてきている。

高圧ガス保安法における国家試験等では、記述式問題は高難易度の試験の一部に短答式問題として採用されており、これらを CBT 方式の試験に置き換えた場合、試験の継続性を担保する必要性から従来と同程度の試験レベルを維持しなければならず、今後さらなる検討が必要である。

²⁷ 石岡 恒憲[2016]”コンピュータ上で実施する記述式試験”，研究報告コンピュータと教育，19，1-2 頁

(4) 問題数

(3)で検討した選択式の問題を主に、これまでの設問を個別に利用していくこととした場合、これまで10問の出題であった科目試験は40問に15問で出題していた科目試験は60問の取扱となる。加えて、IRTでは、新問の出題条件として問題の事前評価が必須であることから、事前評価を目的とした評価問題5問ほどをこれに加えた構成としたい。ここで、評価問題は試験問題と区別のつかない形で出題をするが、試験の得点としては採点を行わない問題となる。

なお、問題の事前評価に必要な受験者データを最低300人分とした仮定した場合、年間1000人の受験者を擁する試験では新問を15問(現行も新問は1科目最大15問)、1万人を擁する試験では新問を166問(現行は新問は各科目30問)評価できる仕組みとなり、新問の供給能力としては現行と遜色のない運用が可能である。

表 19 選択式問題の問題数の見直し案の一例

現在	見直し後(案)
中問(設問4つから構成)10問	設問40問 + 評価問題5問
中問(設問4つから構成)15問	設問60問 + 評価問題5問

(5) 出題問題の選定基準

試験には受験者の能力を測定する性質のものと、ある水準に達しているかどうかの合否判定を行う性質のものがある。前者は学校教育や民間の能力検定において行われることが多く、比較的優しい問題から難しい問題まで織り交ぜて試験を構成し、どの問題まで解くことができるかで理解の到達度を測定している。一方で選抜試験や資格試験では、一定の能力水準に達しているかどうかで合否を出すことを目的としており、合否水準の判定に影響しない問題を出題することに意味はない。しかしながら、受験者の能力を推定した後、合否の水準に照らすという方法も試験としては非効率ではあるが、学力試験に慣れ親しんだ受験者にとっては納得感の得られる方法であると考えられる。

受験者の能力を測定することに関して CBT 方式ではコンピュータ適応型試験²⁸という方法を取ることができる。これは、受験者の能力に応じて適切な問題を選択して出題する方法であり、視力検査を例にとると、視力 0.5 を判別するランドルト環の向きが分かれば、次に出题する問題は視力 1.0 を測定するものとし、これが分からなければ、次に出题する問題は視力 0.7 を測定する問題とするという形で試験を進行し、最終的にどこで安定するかを確認して受験者の視力と推定する手法である。この場合、視力 2.0 が連続で見えていればそこで試験は終了し、結果が安定しなければいつまでも測定が続くということになるため、試験の出题数が受験者によって異なるという特徴がある。

²⁸ Computer Adaptive Test, CAT。

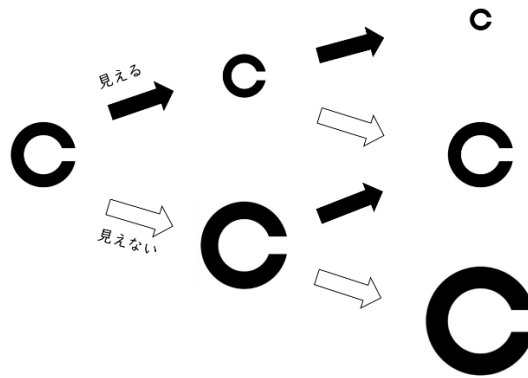


図 5 CAT のイメージ（視力検査）

一方で、可否を判定することに重きをおいた場合、視力 1.0 以上という合格水準の判定をするために視力 0.5 を判別する問題を出題する意味が無い。高圧ガス保安法における国家試験は、可否判定型の試験であるため、出題の効率を考えた場合、合格水準を判別する問題を厳選しこれを出題することが望ましいと考えるが、この場合、合格水準付近の能力を高精度で判別する試験となり、逆に能力が高い者、低い者の間での判別は難しくなる。試験の結果も合格、不合格の通知のみとなるため受験者の試験結果に対する納得感は失われる可能性が高い。

そこで、高圧ガス保安法における国家試験等では、合理的な試験の追求に拘らずに、受験者の能力を測定し、これを合格基準と照らすことで可否を判別することで納得感のある試験運営も視野に置いて引き続き検討することが必要である。

（6）問題評価

出題した問題がどのような受験者に対して最も有効に機能するかどうかを調べることは、試験の妥当性の向上に繋がる。古典的試験理論でも項目反応理論でも一般には、正答率、識別値というデータを収集することとなるが、前者の理論では受験者に対して問題が悪問ではなかったかという事後的な評価以上の活用が難しい。一方で後者の理論では、事前評価した問題を試験に使用できるという利点があり、受験者の解答結果から受験者の能力を推定（＝その能力である確率）することができる。また、異なる試験間の試験の難易度を調整することができ、常に同じ水準で受験者を評価することができることは非 PBT 方式での試験を想定した際には大きな利点と考えられる。

したがって、高圧ガス保安法の国家試験等の非 PBT 化にあたっては、事前評価を行った問題で試験を構成することが望ましい。

なお、問題評価の際の指標は、どの程度能力の受験者を振り分けられるか（識別力）、問題の正答率（困難度）、勘で解いていないか（当て推量）の 3 パラメータが考えられるが、3 パラメータで問題評価を行う場合には、必要な試行回数が 1000 回を超え、また結果の安定性も低下する。一般的に実用の点から当て推量を考慮しない 2 パラメータで評価する事が多く、この場合は試行回数 250 回程度で運用できるようになると考えられている。

(参考) 事前評価

項目特性曲線²⁹ 3パラメータの場合の関数。グラフは縦軸に正答割合、横軸に能力をとる。当て推量が存在するとした場合、能力が低いほど曲線の高さは当て推量に近づく（解答する能力が無く正答数0の人が当て推量の確率で正答となるため）。また、当て推量が大きいほど、能力が低い段階から正答確率が大きくなる。

$$P(u_j=1|\theta) = \frac{1}{1+\exp\{-1.7a_j(\theta-b_j)\}} + c_j \left(1 - \frac{1}{1+\exp\{-1.7a_j(\theta-b_j)\}} \right)$$

a: 識別力 b: 困難度 c: 当て推量 abc を項目パラメータといい、受験者の正誤情報から推定される。c は選択肢の個数の逆数に似る（選択肢2つの場合 1/2）。

$$\theta = b_j \text{ の際の勾配が最も急で、接線の傾きは } \frac{1.7a_j(1-c_j)}{4}$$

(7) 採点

採点は受験者の解答結果に影響を与えるものであってはならない。一方で、(3) で検討したとおり、記述式の問題では採点に偶然誤差が生まれることがある。したがって、採点の面からは客観的に行える問題形式とすることが望ましい。

しかし、客観的な採点においても、考え方は一つではなく、古典的試験理論では、予め問題の難易度を想定して配点を付し、正解した問題の配点を足し上げることで受験者の得点とする事に対し、IRT では、試験結果から受験者の能力を推定（＝その能力である確率）することになる。試験結果から受験者の能力を推定するという行為は、時間逆行の検証であるため本来単純な事象であっても算出が難しい。これを解決する代表的な方法が最尤推定法と Bayes 推定法であるが、最尤推定法は計算が簡易であるが、推定がうまく機能しない条件があるため実用的ではない。一方、Bayes 推定法は最尤推定法の欠点を克服しており、数学的な裏付けが取れた推定が可能である。したがって、IRT を用いる試験では Bayes 推定法を用いて、受験者の能力を推定するという方法で採点を行う。なお、IRT で受験者の能力評価を行う場合、得点という考え方がなくなるが、受験者にとって得点は馴染み深い指標であるため、得点換算式を用いて変換して示すことは、受験者に対する透明性を確保し納得感を創出することができ、有益と考える。

なお、TOEIC では、IRT を用いた測定結果を点数に変換しており、全問不正解でも 10 点、数問間違えても満点を取ることができる。これは、図 6 得点換算式のイメージのような試験特性曲線に対して、これを直線に近似する得点換算式を用いているためと考えられ、ある範囲より大きい又は小さい能力の者に対しては、与えられる得点が固定となるためである。この方法は計算が簡易になるという利点があるが、この近似を行わずとも計算は十分可能であるため、特段の支障がなければ近似を行わず得点換算を行うことが望ましい。

²⁹ 高橋信「IRT 項目反応理論入門 統計学の基礎から学ぶ良質なテストの作り方」（オーム社、2021 年）50 頁

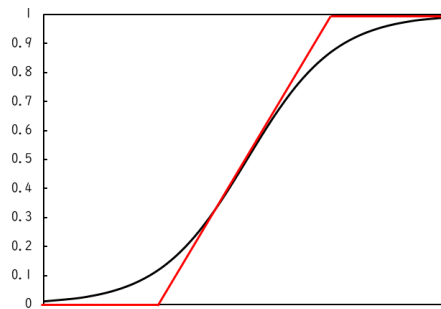


図 6 得点換算式のイメージ

(8) 合否判定基準

非 PBT 化によって試験の実施回数が増加した場合、試験の公平性を保つために、合否判定基準の設定についても検討が必要である。

具体的には、資格試験においては、ある一定以上の能力レベルを超えた受験者全員に対して「合格」という評価をするが、この合否判断の基準となる「一定の能力レベル」の線引き（カットオフポイント）をどのように決定するかが問題となる。特に、複数回実施する試験の場合、どの実施回においても「ある一定の能力レベル＝合否基準」であることが求められる。そのためには、まず規準集団となるモニター受験者に対して予備テストを行い、規準集団上で合否を分けるカットオフポイントを定めた上で、複数回のテストの能力尺度を共通尺度化（等化³⁰）する。この手法により、カットオフポイントの決定を毎回の本試験を行うたびにを行う必要がなくなり、効率的なテストの実施につながる。また、カットオフポイントの設定について、本試験の実施とは別にテスト実施者が検討することができ、より公平なテストの実施が期待できる。

高圧ガス保安法における国家試験等においても、非 PBT 方式による場合は実施方式、出題方式の変更が想定されるが、現行の国家試験等と合否判定基準が変わらないようにするためには事前にモニター受験者に対して予備テスト等を行い、評価することが重要である。

(9) 試験時間

(4) の検討において、CBT 方式と IRT を活用した試験では選択式問題の問題数が 1 割ほど増加する見込みとなった。問題数の増加に合わせて試験時間を見直すことを考えた場合、60 分試験では 6 分、120 分試験では 12 分の試験時間の拡大が必要となり、1 問当たりの解答時間にすると 12 秒ほどの短縮に相当する。しかしながら、PBT 方式にて必要であったマークシートへの転記及びその見直しの時間が CBT 方式では不要になるため、試験時間の延長はこれと相殺し、行わなくても影響は小さいと考える。

(10) 試験回数（受験機会）

高圧ガス保安法における国家試験は、同法により毎年少なくとも 1 回実施することとされ、現

³⁰ 「等化」(equating)：異なるテスト得点尺度を、共通の尺度上で表現する操作を呼ぶ。等化を行うためには、異なるテストがどれも同じ概念を測っており、妥当性が高いことが要求されている。

に年に1回、実施している。高圧ガス保安法に係る免状は、国家試験に合格しなければ、その交付を受けることができないが、個人の理由や特別な社会情勢等の影響により受験機会を失ってしまった場合、1年間、免状取得を希望する者は保安活動に携わることができなくなる。

本事業で実施したアンケート調査及びヒアリング調査の結果からも、各試験実施団体は、受験者の受験機会の拡大を目的として非PBT方式を採用することがわかっている。十分な問題量が確保できている場合には、非PBT方式によって、特定の試験日ではなく、ある一定程度の実施期間（例：通年、1か月等）を設定し、1年間で実施する回数を増やすことが可能であると考えられる。

（II）問題の公開

試験問題の公開は、試験対策に効果的な教材として受験者に活用されている。政府方針としても「資格取得の容易化」を目的として、1999年と2003年に「試験問題の公表・持ち帰りの推進」に関する閣議決定³¹³²がされ、2011年の総務省勧告³³においても、すべての国家試験に対して、「過去の試験問題やその解答について、インターネット等により積極的に無償で公開すること」を求めている。

一方で、問題を公開していない国家試験もあり、こうした国家試験に対しては、情報公開法³⁴に基づく開示請求が行われ、当該機関が不開示の決定をした際には、情報公開・個人情報保護審査会に対して不服申立てがされている例もある。この結果、開示の判断が確定したものは、医師国家試験と司法書士資格認定試験の事例がある。これらについては、若林氏の事例研究論文³⁵に詳しく説明されており、同氏は公開の判断のポイントを次のように整理している。

- ① 試験問題公開による透明性確保の必要度
- ② 試験問題の再利用度
- ③ 試験問題を活用した試験対策の許容度

³¹ 規制緩和推進3か年計画（改定）（平成11年3月30日閣議決定）.pdf P6。

（II）例えば以下の方法を採用することにより、資格取得の容易化を検討する。

合格科目の積み上げ方式による合格方式の推進

再受験における既合格科目の免除制度の推進

試験問題の公表・持ち帰りの推進

³² 規制改革推進3か年計画（再改定）（平成15年3月28日閣議決定）.pdf P14

³³ 検査検定、資格認定等に係る利用者の負担軽減に関する調査結果に基づく勧告（2011）.pdf P26

3 申請手続の負担軽減等の推進

【所見】

（3）利用者への配慮

② 手数料等の内訳・割引条件・割引額、資格取得要件など制度を利用するに当たり必要となる基本的な情報を的確に利用者に提供すること。（厚生労働省、経済産業省、国土交通省）

また、過去の試験問題やその解答について、インターネット等により積極的に無償で公開すること。（総務省、文部科学省、厚生労働省、経済産業省、国土交通省、環境省）

³⁴ ここでは「行政機関の保有する情報の公開に関する法律」と「独立行政法人等の保有する情報の公開に関する法律」の総称として呼称する。

³⁵ 若林 昌子[2015]，「わが国の公的試験における試験問題公開の判断基準－情報公開制度における事例－」，日本テスト学会誌

④ 試験問題開発の労力増加の受容度

⑤ 試験問題の情報管理度

このうち、①が政府方針による受験者への配慮、②-⑤は試験の運営上の観点であり、試験の本質に影響を与える要素となる。司法書士資格認定試験では①を基に、公的な試験の問題は公表して透明性を確保すべきとの判断があった。

②の問題の再利用については、公務員採用Ⅲ種試験では、公平性を確保するため、複数の試験について内容やレベルの標準化が図れるよう同じ試験問題を繰り返し出題しているという例もある。この場合は問題の公開は求められていないが、司法書士資格認定試験の例では、すべてがそのまま用いられるわけではない場合は、同じ試験問題が出題されているとは解さないとされている。

③の過去問による試験対策については、司法書士資格認定試験では「単なる傾向にとどまらず内容についても相当程度に推測される」としているが、推測可能性や解答の画一化は部分的と判断されている。また、医師国家試験においても、過去の試験問題が出題された場合に正答率が向上するとは認められないとの判断から、過去の試験問題を通じて知識を習得することを問題視していないため公開の判断に至っている。

④の作問の労力については、司法書士資格認定試験では、試験問題を公開した場合、試験問題開発の労力が増加するとの指摘に対して、国が開発した試験問題である場合、その負担は受忍すべきとされている。一方で、防衛庁（防衛省）の職員採用試験では作問を民間事業者が行っており、一般に公開しない契約であることから不開示となっている（ただし、公開を前提とした契約に見直すよう言及されている）。

⑤の問題の情報管理については、医師国家試験のように事実上明らかにされてしまっている場合には公開すべきものと判断されている。試験問題を能力評価試験のノウハウと扱う場合には、知的財産、秘密情報としての管理状態が審議されている。

ここで、上記の公開に関しての論点中、高圧ガス保安法においては、①、⑤は、すでに問題を公表しているため、公開すべきとの判断になるものと考えられ、②、③の過去問に関しては、協会は過去問題集を試験で良い点数をとるための教材として扱っている事実がある。また、過去問の利用は、系統誤差を低下させ、妥当性の向上にも寄与する³⁶。

過去問を教材とすることへの逆の評価としては、テストワイズネスと呼ばれる本来の能力より高い点数をとる者の存在も指摘されており、この集団は、テスト形式が変わると得点が取れなくなる等の特徴を持つ。また、過去問の利用による公平性については、試験対策の公平性は確保されるが、経済的理由による過去問へのアクセスの不公平性も指摘される。

④の作問の労力増加に関しては、現行は全て新問で運用しており、今後のIRTを用いた試験の想定においても、IRTではそもそも新問で受験者の能力を評価することはできないため、原則過去問を利用することとなり、現状以上の新問作成の労力を要しない。

これらを総合すると、高圧ガス保安法における今後の試験においては、非PBT方式で実施する

³⁶ 馬場 正太郎[2021], ‘過去問の定義と公平性・妥当性との関連-日本の大学入試対策用の英語教材を分析対象として-’ 日本テスト学会誌, 77-78

ときは、試験問題を公開することは難しいことから、試験のレベルを推し量れるような参考問題等の公開を行うことが望ましい。ただし、問題データベースの総量が十分でない期間の公開は、試験の妥当性に影響を与えるため、当分の間は非公開とする一時的な措置を講じる必要がある。

(12) 作問

現在、国家試験の作問は、毎年、外部専門家等による委員会で審議し、作成している。これを IRT を用いた非 P B T 方式にて実施する場合、問題は原則問題データベースから出題されるため、作問作業はデータベースの整備が主となる。一方で、新問については、現在と同様の方法で問題を作成するが、IRT による事前評価を行い、数学的裏付けに基づく問題の評価を行った問題を問題データベースに登録する方法が望ましい。

(13) 受験会場

テストの公平性の確保をする上では、受験環境の不均一性の排除が挙げられる。試験の実施方式に関わらず受験環境の基本的な条件として、適切な冷暖房や照明の提供、騒音の防止等がある。

非 PBT 方式の試験においては、これらの他に試験において使用する機器（PC 等）やネットワーク環境の設定、その他の必要な機材（例：キーボード、マイク、ヘッドフォン等）の仕様を適切に定めて明示し、全ての受験者が同一の環境下で受験できるよう会場を整える必要がある。

CBT 方式は集合型であるため、受験環境の基本的な条件に加え、会場で既に整備している機器及びネットワーク環境等についても試験実施団体（又は非 PBT 事業者）によって統一的に管理することができるため、従来の PBT 方式での試験運営方法との類似性が高く、受験環境の統制が可能であると考えられる。また、PBT 方式の全国一斉の試験実施という制約から開放されるため、その利点を活かし、全国各地の小規模会場（テストセンター）を活用し、一定期間内での分散開催として受験機会の拡大を図ることが望ましい。なお、会場の設定に際しては、離島などでは試験受験に係る経済的、心身的な負担が他の地域と比較して大きいことが指摘されており、受験者の負担ができる限り少なくなるよう会場の設定することで、経済的負担の軽減に配慮することが望ましいと考える。

4. 2 非 PBT 方式で国家試験等を実施する場合の制度的制約

高圧ガス保安法では、製造保安責任者試験及び販売主任者試験について、以下のとおり規定されている。

高圧ガス保安法（昭和二十六年法律第二百四号）

（製造保安責任者試験及び販売主任者試験）

第三十一条 製造保安責任者試験又は販売主任者試験は、高圧ガスの製造又は販売及び高圧ガスによる災害の発生の防止に関して必要な知識及び技能について行う。

2 製造保安責任者試験又は販売主任者試験は、第二十九条第一項に規定する製造保安責任者免状又は販売主任者免状の種類ごとに、毎年少なくとも一回、経済産業大臣又は都道府県知事が行う。

3 協会又は経済産業大臣が指定する者（以下「指定講習機関」という。）が経済産業省令で定めるところにより行う講習の課程を修了した者については、経済産業省令で定めるところにより、製造保安責任者試験又は販売主任者試験の全部又は一部を免除する。

4 前三項に定めるもののほか、製造保安責任者試験又は販売主任者試験の試験科目、受験手続その他の細目及び前項の指定に関し必要な事項は、経済産業省令で定める。

（備考）上記で「協会」とあるのは、高圧ガス保安協会を指す。（以下、本項において同じ。）

以上のとおり、第31条第1項と第2項で製造保安責任者試験と販売主任者試験に関する内容と回数について、同条第3項で協会又は指定講習機関が実施する講習について定められている。

さらに、協会又は指定講習機関が実施する講習の詳細については、高圧ガス保安法に基づく高圧ガス製造保安責任者試験等に関する規則（以下「試験則」という。）第4条および第5条に規定されている。各条文を以下に示す。

高圧ガス保安法に基づく高圧ガス製造保安責任者試験等に関する規則（昭和四十一年通商産業省令第五十四号）

（協会又は指定講習機関が行う講習の方法）

第四条 法第三十一条第三項の協会又は指定講習機関が行う講習は、次の表の講習の種類に掲げる講習の種類に応じて、それぞれ同表の講習科目の欄に掲げる科目について行い、かつ、各科目ごとに七時間以上行わなければならない。

講習の種類	講習科目		
甲種化学講習	法に係る法令	高圧ガスの製造（冷凍のための製造を除く。以下この項から丙種化学特別講習の項までにおいて同じ。）に必要な化学に関する高度の保安管理の技術	高圧ガスの製造に必要な高度の応用化学
甲種機械講習		高圧ガスの製造に必要な機械に関する高度の保安管理の技術	高圧ガスの製造に必要な高度の機械工学

乙種化学 講習	同右	高压ガスの製造に必要な化学に 関する通常の保安管理の技術	高压ガスの製造に必要な 通常応用化学
乙種機械 講習		高压ガスの製造に必要な機械に 関する通常の保安管理の技術	高压ガスの製造に必要な 通常機械工学
丙種化学 液石講習	同右	液化石油ガスの製造に必要な通 常の保安管理の技術	液化石油ガスの製造に必 要な通常応用化学及び 基礎的な機械工学
丙種化学 特別講習	同右	高压ガスの製造に必要な基礎的 な保安管理の技術	高压ガスの製造に必要な 基礎的な応用化学及び基 礎的な機械工学
第一種 冷凍機械 講習	同右	冷凍のための高压ガスの製造に 必要な高度の保安管理の技術	冷凍のための高压ガスの 製造に必要な通常応用 化学及び機械工学
第二種 冷凍機械 講習	同右	冷凍のための高压ガスの製造に 必要な通常保安管理の技術	冷凍のための高压ガスの 製造に必要な基礎的な応 用化学及び機械工学
第三種 冷凍機械 講習	同右	冷凍のための高压ガスの製造に 必要な初歩的な保安管理の技術	
第一種 販売講習	同右	高压ガス（液化石油ガスを除 く。）の販売に必要な通常保安 管理の技術	
第二種 販売講習	法に係る法令 及び液化石油 ガスの保安の 確保及び取引 の適正化に関 する法律（昭 和四十二年法 律第百四十九 号。以下「液化 石油ガス法」 という。）に係 る法令	液化石油ガスの販売に必要な通 常の保安管理の技術	

（第2項～第4項 略）

（備考）上記で「法」とあるのは、高压ガス保安法を指す。（以下、本項において同じ。）

（協会又は指定講習機関が行う技術検定）

第五条 協会又は指定講習機関は、前条の規定による講習を受けた者に対して、その講習に係る
 高压ガスの製造又は販売に必要な保安管理の技術及び高压ガスの製造に必要な応用化学又は
 機械工学について技術検定を行わなければならない。

また、製造保安責任者試験と販売主任者試験に関する試験科目等については、試験則第9条に
 規定されている。条文を以下に示す。

(試験科目等)

第九条 製造保安責任者試験及び販売主任者試験は、筆記による学科試験とし、その試験科目は、
 次の表の製造保安責任者免状及び販売主任者免状の種類の欄に掲げる製造保安責任者免状及
 び販売主任者免状の種類に応じて、それぞれ同表の試験科目の欄に掲げるものとする。

製造保安責任者免状及び販売主任者免状の種類	試験科目		
甲種化学責任者免状	法に係る法令	高压ガスの製造（冷凍のための製造を除く。以下この項から丙種化学特別講習の項までにおいて同じ。）に必要な化学に関する高度の保安管理の技術	高压ガスの製造に必要な高度の応用化学
甲種機械責任者免状		高压ガスの製造に必要な機械に関する高度の保安管理の技術	高压ガスの製造に必要な高度の機械工学
乙種化学責任者免状	同右	高压ガスの製造に必要な化学に関する通常の保安管理の技術	高压ガスの製造に必要な通常の応用化学
乙種機械責任者免状		高压ガスの製造に必要な機械に関する通常の保安管理の技術	高压ガスの製造に必要な通常の機械工学
丙種化学責任者免状	同右	液化石油ガスの製造に必要な通常の保安管理の技術	液化石油ガスの製造に必要な通常の応用化学及び基礎的な機械工学
第一種冷凍機械責任者免状	同右	冷凍のための高压ガスの製造に必要な高度の保安管理の技術	冷凍のための高压ガスの製造に必要な通常の応用化学及び機械工学
第二種冷凍機械責任者免状	同右	冷凍のための高压ガスの製造に必要な通常の保安管理の技術	冷凍のための高压ガスの製造に必要な基礎的な応用化学及び機械工学
第三種冷凍機械責任者免状	同右	冷凍のための高压ガスの製造に必要な初歩的な保安管理の技術	

第一種販売主任者免状	同右	高圧ガス（液化石油ガスを除く。）の販売に必要な通常の保安管理の技術	
第二種販売主任者免状	法に係る法令及び液化石油ガス法に係る法令	液化石油ガスの販売に必要な通常の保安管理の技術	

2 丙種化学責任者免状に係る製造保安責任者試験を受けようとする者は、前項の規定にかかわらず、同項の表丙種化学責任者免状の項試験科目の欄に掲げる試験科目に代えて法に係る法令、高圧ガスの製造（冷凍のための製造を除く。以下この項において同じ。）に必要な基礎的な保安管理の技術並びに高圧ガスの製造に必要な基礎的な応用化学及び基礎的な機械工学（以下「特別試験科目」という。）を試験科目とする旨の申請をすることができる。
（第3項～第4項 略）

これら示した条文のうち、国家試験等に非 PBT 方式を導入するにあたっては、試験則第9条において規定されている『製造保安責任者試験及び販売主任者試験は、筆記による学科試験』の文言を改める必要がある。

既に非 PBT 試験を導入している情報処理技術者試験では、情報処理の促進に関する法律施行規則において、以下のように規定されている。

情報処理の促進に関する法律施行規則（平成二十八年経済産業省令第百二号）

第三十七条 （第1項 略）

2 技術者試験は、筆記試験又は電子計算機その他の機器を使用して行う試験により行うものとする。

これを参考に、国家試験等に非 PBT 方式を導入するにあたり、試験則第9条の改正案文を以下のとおり示す。

高圧ガス保安法に基づく高圧ガス製造保安責任者試験等に関する規則

改正前	改正後
（試験科目等）	（試験科目等）
第九条 製造保安責任者試験及び販売主任者試験は、筆記による学科試験とし、その試験科目は、次の表の製造保安責任者免状及び販売主任者免状の種類の欄に掲げる製造保安責任者免状及び販売主任者	第九条 製造保安責任者試験及び販売主任者試験は、筆記による学科試験又は電子計算機その他の機器を使用して行う学科試験とし、その試験科目は、次の表の製造保安責任者免状及び販売主任者免状の種類の欄

免状の種類に応じて、それぞれ同表の試験科目の欄に掲げるものとする。 2～4 略	に掲げる製造保安責任者免状及び販売主任者免状の種類に応じて、それぞれ同表の試験科目の欄に掲げるものとする。 2～4 略
--	--

4. 3 国家試験等実施期間等の第三者による受験者本人の確認方法

CBT 方式による試験における本人確認は、アンケート調査等の結果を踏まえ、会場受付で写真付の公的身分証明書と来場した本人を確認する方法又は事前提出させた写真と写真無の公的身分証明書のセットに加え、来場した本人を確認する方法が、本人性の確認と同一性の確認を満たすことができるものと考えられる。

IBT 方式による試験においても同様に、事前に登録した写真付の公的身分証明書とカメラを用いて本人の確認及び監視の方法又は事前に登録した写真と事前に登録した写真無の公的身分証明書のセットに加えカメラを用いて本人の確認及び監視の方法が適切と考えられる。

ただし、令和3年度報告書では、IBT 方式の試験では不正行為防止措置に係る技術発展が不十分であり、替え玉受験を含む不正行為を適切に判定できない可能性を指摘したところである。

したがって、本人確認については、方法に加えて実際に機能させることができるのかという面での考慮も行う必要がある。現時点では、上記の理由により、CBT 方式における本人確認は会場受付で写真付の公的身分証明書と来場した本人の確認又は事前提出させた写真と写真無の公的身分証明書のセットに加え、来場した本人の確認による方法が適切と考える。

4. 4 試験実施団体における国家試験等の非 PBT 化の有用性及び非 PBT 事業者の選定時のポイント

誰の立場から考えるのかによって、試験の有用性の程度は変わり得るため、本項においては、試験実施団体の立場に限定して、試験を非 PBT 化する有用性について整理した。また、試験実施団体が、非 PBT 方式への移行に取り組むためには、試験システム及び試験運営サービスの委託先として非 PBT 事業者を選定しなければならないため、選定時に検討すべき主なポイントについても整理した。

(1) 試験実施団体における国家試験等の非 PBT 化の有用性

イ. コスト削減

本事業で実施したヒアリング調査の結果から、複数の試験実施団体が、コロナ禍による試験会場や試験監督官に係る費用の増加がきっかけとなり、PBT 方式から非 PBT 方式への移行に取り組んだことがわかった。また、今後、わが国においては、少子高齢化による総人口と生産年齢人口の更なる減少によって、深刻な働き手不足が続くと予想されている。試験実施団体は、非 PBT 化によって試験実施に係るコスト（労力）を削減することができる。具体的には、PBT 方式の試験を実施するためには、表 17 に示したとおり、一般的に会場の手配や賃借料だけでなく、試験用紙の印刷・配送等、試験事務局において様々な業

務が発生する。しかし、試験を非 PBT 化することによって、PBT 方式での試験実施に必要な会場探し、会場設営、申込・受付、試験監督、結果通知等に係る労力については、システム化される、又は非 PBT 業者により試験運営サービスの一環として対応されるため、試験実施団体に係る全体的な労力は低減されるであろう。

ただし、非 PBT 化によって、試験実施に係るコスト（費用・労力）が必ず削減できるわけではなく、試験の規模が重要な判断要素の一つとなる。今回ヒアリングした非 PBT 事業者からは、非 PBT 化に係るコストについて正確な金額は得られなかったが、試験の規模が大きければ（例：受験者数が年間 10 万人程度）CBT 方式に移行することでコストが抑えられ、受験者数の少ない試験（例：年間 100 人程度）は費用対効果という観点から非 PBT 化は割に合わないという見解が示された。

ロ．人的リスクの低減

上記で述べた非 PBT 化による試験実施に係る労力の低減は、PBT 方式よりも試験の実施に関わる人を限定できるということも意味しており、人的リスクの低減を図ることができると考えられる。

本事業のヒアリング調査の結果から、非 PBT 方式に移行した後も、試験問題については、PBT 方式と同様に、外部の専門家等によって構成される作問委員会によって作成されていることがわかっている。しかし、非 PBT 化することで試験問題の印刷・保管や会場への配送等の作業工数が減り、作業に関与する従業員や外部業者を絞り込むことができるため、問題漏洩が発生する可能性を低減することができる。

また、試験問題及び解答用紙の取扱いについては、試験制度の公平性や社会的信頼性等の観点から様々な注意が必要であるが、非 PBT 化によって、試験問題及び解答用紙の誤配送等が発生する可能性も低減することができる。

なお、非 PBT 化によって、このような人的リスクの低減は可能であるが、同時に、非 PBT 事業者はセキュリティに最大限考慮したテスト体制の整備が必要であるため、後述する非 PBT 事業者の選定時にコンピュータ技術者やネットワーク技術者等の専門家の配置についても考慮すべきである（これら専門家のコストについては、4. 4（2）ロを参照されたい。）。

（2）非 PBT 事業者の選定

イ．非 PBT 化への理解及び実績

試験の非 PBT 化に携わる非 PBT 事業者には、試験システムの構築・運用だけでなく、非 PBT 化に適した作問の支援や試験運営サービス等の提供が求められる。PBT 方式で実施していた試験を CBT 化する場合、試験実施団体が非 PBT 事業者へ委託する業務については、表 17 において整理したとおり多岐にわたる。加えて、非 PBT 事業者は、これらの業務を行う上で、「テスト・スタンダード」において定められている、CBT 試験の公平性を保つために必要な技術的要件（使用端末や実施環境の仕様、ユーザインタフェースの在り方、不測

の事態への対応等)についても満たすことが求められる。

非 PBT 事業者は、これらのことを理解し、求められる業務を遂行するために、コンピュータ技術者、ネットワーク技術者、計量心理学者、テストの企画・作問及び設計実施管理者(テストディベロッパー)等、広範囲の分野に関する専門家の知識、ノウハウ、経験を集結させて取り組む必要がある。また、これらの専門家を配置するだけではなく、非 PBT 事業者において前述の業務を実際に遂行した実績を有していることも肝要である。

本事業で実施したヒアリング調査の結果から、非 PBT 事業者は、非 PBT 試験事業を展開する上で、自社の強みとして厳格性と利便性のいずれかを重視している傾向が見てとれた。事業者によっては、CBT 方式の試験についての実績はあるが、IBT 方式については実績がない等のケースが想定されるため、委託前に実績について確認することが重要である。

特に、厳格性が求められるハイ・ステークスな試験(国家試験など)においては、本人確認、受験環境の統制、試験監督等の不正行為を防止するための対策について、試験システムと試験運営の両面から検討するとともに、試験実施団体が採用したいと考えている非 PBT 方式での実績を確認した上で、非 PBT 事業者を選定する必要がある。

実際、本事業で実施したヒアリング調査では、非 PBT 方式の試験における本人確認、受験環境の統制、試験監督等の不正行為防止の対策については、非 PBT 事業者の定める基準やルールに基づき実施していると回答する試験実施団体が複数存在した。このことを踏まえると、非 PBT 事業者の実績をはかる具体的なポイントとして、不正行為防止の対策について、どのような方針とノウハウを持ち、どの程度確立されたマニュアルが策定されているのかを確認することが有用と考える。

ロ. 非 PBT 化への理解及び実績非 PBT 化に必要な費用

表 17 で示した CBT 方式で試験を実施する場合の費用項目を踏まえると、非 PBT 化に伴い新たに追加される費用項目は、試験システム(申込受付システム、決済システム、テストシステム、受験者データシステム)の導入・維持・管理、試験の編集(問題のデータ化、試験問題の質担保を目的としたコンサルテーション)である。試験実施団体は、自組織において既に保有している試験システムの状況や試験編集のノウハウ等を踏まえた上で、非 PBT 事業者に対する新規費用項目(全て又は一部)の委託の可否を検討する必要がある。また、非 PBT 方式で試験を実施するためには、試験システムのセキュリティ対策が必須であるため、試験システムの導入・維持・管理に関する費用項目にそのためのコストが含まれているか否かも確認するとよい。

なお、本事業のヒアリング調査の結果から、非 PBT 事業者は、試験の非 PBT 化に必要な作業、コスト(費用と労力)、顧客のニーズ等を考慮して多様なサービスや契約プランを用意していることがわかっている。試験実施団体は、非 PBT 事業者に対する委託要件を検討する際、試験の申込み～受験料決済～試験配信～結果通知という一連の流れの中で、非 PBT 化によって誰(例:受験者、社会一般、試験実施団体等)のためにどのような試験の厳格性、利便性及び有用性を求めるのかを整理し、その上で自組織の実施している試験の現状や今後の展望(例:年間受験者数、試験実施回数・会場等、独自試験システムの運用等)

を踏まえ、非 PBT 事業者に対する適切な委託要件を詰めていくとよい。これにより、試験実施団体において非 PBT 方式への移行に必要な費用についても明確化できるであろう。

ハ. 非 PBT 事業者が満たすべき最低限の要件

4. 4 (2) ロにて述べたとおり、非 PBT 事業者に対する委託要件については、試験実施団体によって何を適切な内容とするかの判断が異なるが、ここでは非 PBT 事業者が満たすべき最低限の要件について、「a)試験の公平性を保つための機能・仕組み」「b)合理的配慮への対応」「c)不測の事態への対応体制」の観点から以下のとおり整理した。

a) 試験の公平性を保つための機能・仕組み

前述しているとおり、試験の厳格性と受験者の利便性はトレードオフ関係にある（4. 1 (1) 参照）。試験実施団体は、試験の特性（ハイ・ステークスか否か、年間受験者数や年間実施回数、等）や非 PBT 化の目的を踏まえ、試験システムと試験運営サービスに対してどの程度の厳格性と利便性を求めるのかを整理し、その上で非 PBT 事業者に対して試験の公平性を保つために必要な機能及び仕組みの有無や内容について確認する必要がある。具体的には、試験システム及び試験運営サービスの両方において、受験環境の統制、受験者の本人確認（申込時・受験当日）、試験監督、試験中のノートテイキングの統制等に関して、非 PBT 事業者が試験実施団体の必要とする機能及び仕組みを有し、またそれらをどのように組み合わせて用いることができるのかを確認するとよい。

b) 合理的配慮への対応

2016 年 4 月から、事業者には、障害者差別解消法³⁷によって障害のある方への「合理的配慮」等が求められている。合理的配慮の提供とは、事業者に対して、障害のある人から社会の中にあるバリアを取り除くために何らかの対応を必要としているとの意思が伝えられた時に、負担が重すぎない範囲で対応に努めることである³⁸。

このことを踏まえると、特に国家試験のようなハイ・ステークスな試験については、試験実施団体が非 PBT 事業者を選定する際に、障害のある方からの要望に応じた試験会場での配慮（例：操作補助者の入室許可）や、試験システム（申込・結果通知システム、試験配信システム）のユーザビリティやアクセシビリティ（例：表示画面の拡大機能、画面に表示されている色の反転機能等）等に関する取組みについても検討の要素に加える。なお、試験システムのアクセシビリティについては、WCAG（Web Content Accessibility Guidelines）³⁹という、様々な障害のある方でも Web コンテンツが利用できることを目指

³⁷ 法律の正式名称は「障害を理由とする差別の解消の推進に関する法律」

³⁸ 出典：内閣府「合理的配慮」を知っていますか？」

https://www8.cao.go.jp/shougai/suishin/pdf/gouriteki_hairyo/print.pdf (2023/3/2 時点)

³⁹ WCAG は、Web コンテンツをよりアクセス可能にするための幅広い推奨事項を定めたガイドラインである。これらの推奨事項に基づいた Web コンテンツを制作することにより、失明や弱視、難聴、運動制限、言語障害、光線

したガイドラインが定められている。非 PBT 事業者を選定する際のポイントとして、同ガイドラインで示されている推奨事項が参考になる。

c) 不測の事態への対応体制

試験を実施する上で、以下のような不測の事態の発生は事前に想定しておくべき事項である。非 PBT 化により試験の受験機会が増加すれば、不測の事態が発生する可能性も高まることとなる。試験実施団体は、非 PBT 事業者を選定する際、不測の事態に対して予めどのような体制で臨むのか、またどのような対応マニュアルが策定されているか等について、確認するとよい。

<想定される不測の事態>

- 悪天候や災害等によって試験が実施／受験できない場合
- 受験者の急病や身内の不幸等、受験者の責めに帰すことができない事情によって受験できない場合
- 受験者によって不正行為（替え玉受験、カンニング、問題漏洩）が行われた場合
- 試験前／試験中に使用端末やネットワーク環境の不具合が発生した場合
- その他、受験者からの問合せがあった場合

二. ベンダーロックインの回避

公正取引委員会によればベンダーロックインとは以下の状態のこといい、官公庁における情報システムの調達についてはベンダーロックインの回避が喫緊の課題となっている。

ソフトウェアの機能改修やバージョンアップ、ハードウェアのメンテナンス等、情報システムを使い続けるために必要な作業を、それを導入した事業者以外が実施することができないために、特定のシステムベンダーを利用し続けなくてはならない状態⁴⁰

国家試験の実施団体においても、試験システムの調達を定期的に行う可能性を考慮し、ベンダーロックインを回避するための方策を考える必要があり、また、本事業のヒアリング調査においても、試験実施団体からベンダーロックインを懸念する声が実際に聞かれた。

公正取引委員会の「官公庁における情報システム調達に関する実態調査報告書（令和 4 年 2 月）」によると、ベンダーロックインを回避して多様なベンダーの参入を促すためには、既存のシステムベンダーしか参入できないような仕様書となるのを避けることや、既

過敏症、学習障害等、幅広い障害を持つ人々が Web コンテンツにアクセスしやすくなる。ただし、これらの障害を持つユーザーの全てのニーズに対応するわけではない。

参考：Web Content Accessibility Guidelines (WCAG) 2.1 (<https://www.w3.org/TR/WCAG21/>) 一部翻訳 (2023/2/14 時点)

⁴⁰ 出典：公正取引委員会(令和 4 年 2 月 8 日)官公庁における情報システム調達に関する実態調査について https://www.jftc.go.jp/houdou/pressrelease/2022/feb/220208_system.html (2023/3/6 時点)

存システムの機能（技術）等に係る権利が既存ベンダーに帰属している状況に陥らないようにすること等の対応があると示されている。また、情報システムを疎結合化して API 連携を確実にを行うためにデータの標準化や情報システムのオープンソース化が有益であると記されている。

特に教育やアセスメントの分野に関しては、国際標準化団体である IEdTech [旧：IMS Global Learning Consortium(IMS グローバル)] が、情報システムのオープンソース化やデータの標準化を目的とした国際技術標準「IEdTech 技術標準」を開発・提供している⁴¹。試験実施団体は、ベンダーロックインを回避するために、このような国際技術標準の採用の有無についても、非 PBT 事業者を選定する際のポイントとして検討することが望まれる。あわせて、試験システムの事業者間移行にあたってはシステム全体のセキュリティを保つことが大前提であることから、国際技術標準をベースとした試験システムの移行実績についても、事業者選定の検討材料となる。

⁴¹ （一社）日本 IMS 協会 <https://www.imsjapan.org/>（2023/3/6 時点）

5 まとめ

本報告書では、現在 PBT 方式にて実施している高圧ガス保安法における国家試験等の実施方法の在り方について、まず(1)「国家試験等の実施方法の他事例調査」を行い、その結果を基に(2)「非 PBT 方式による国家試験等の在り方の検討」のため、①「非 PBT 方式で国家試験等を実施する場合の実施内容及び制度的制約」、②「国家試験等実施機関等の第三者による受験者本人の確認方法」に関し、現行の PBT 方式との連続性・整合性を確保しながら、非 PBT への転の可能性について考察を行った。以下に、検討の結果を総括する。

(1)「国家試験等の実施方法の他事例調査」のまとめ

本調査にて試験実施する機関・団体・学校 67 団体へアンケート調査依頼の後、非 PBT 方式（国家試験）3ヶ所、非 PBT 方式（民間試験）3ヶ所、PBT 方式（国家試験）4ヶ所及び非 PBT 事業者4ヶ所へヒアリングを実施した。アンケート調査の結果、40 団体の回答が得られ、うち非 PBT 方式にて試験を実施している団体は、国家資格試験関係は4団体、民間資格試験関係は6団体の計10団体（25%）であった。コロナ禍においても PBT 方式での実施事例が多くを占めていた。

非 PBT 方式を採用した理由として、「受験者の実施利便性の向上」「受験機会の拡大」が主であったことから、受験者の立場を考えた実施方針であることが伺えた。また、非 PBT 方式の試験実施方法については、CBT 方式と IBT 方式それぞれの方式があるものの、CBT 方式での実施をしている試験実施団体がほとんどであった。

試験実施団体が非 PBT 方式で実施する際、課題として挙げられたのは、試験システム及び試験の運営を外部の非 PBT 事業者へ委託しているのが過半数を占めていることもあり「導入コストが高額」という点と受験機会の増加に伴う「問題作成数の不足」が多く挙げられた。

前者の「導入コストが高額」については、試験ごとに受験料や受講者数等の要素が異なっていること、また、各試験実施団体から具体的なコストバランスの提示がなかったことから、その程度を計り知れなかった。一方で、PBT 方式にて実施する際の労力（会場探し、会場設営、申込・受付、試験監督、結果通知等）をシステム化することによるコスト減や非 PBT 化することで様々な分析が可能になるため、試験の価値・質の向上を図ることができることなどの、非 PBT 化の有益性が高いことが分かった。

また、後者の課題である「問題作成数の不足」については、人的リソースの観点での課題でもあるため明確な解決策は確認できなかったものの、問題のデータベース等の活用や非 PBT 事業者による作問の専門家（テストディベロッパー）の活用等により問題の質を上げることができていることが分かった。

PBT 方式で実施している試験実施団体の非 PBT 方式の不採用理由として「本人確認が困難」と回答した試験実施団体は多かったが、非 PBT 方式の導入時の課題として回答した試験実施団体はなかったため、非 PBT 方式での本人確認について認識にギャップがあることが分かった。

(2) 「非 PBT 方式による国家試験等の在り方の検討」のまとめ

以上の状況を踏まえて法における国家試験等の在り方を検討した結果、現状（PBT 方式）と CBT 方式（非 PBT 方式）での実施方法。内容として次のようなものが想定される。なお、本人確認の厳格性及び不正防止の観点から CBT 方式での実施を想定してとりまとめた。

① 「非 PBT 方式で国家試験等を実施する場合の実施内容及び制度的制約」のまとめ

A 出題方式

(1) 現状（PBT 方式）

現在実施している国家試験等は、同一の種類資格に関する試験（例 乙種機械）であれば、同一内容の問題を出題することで不公平の発生を抑止している。

上記を担保するため、試験日時を同時にし、全国一斉に実施している。

(2) CBT 方式で実施する場合の論点（非 PBT 方式）

a) 実施方法（案）

問題データベースからコンピュータが出題する問題を、同等程度の難易度の問題グループからランダムで抽出することが可能であるため、出題内容の不公平の発生を防止するだけでなく、試験直前まで問題内容が確定しないので、問題流出の防止に寄与する。

運営面では、試験問題の印刷、発送、採点などの作業が削減され、合理的な実施が可能である。

b) 課題

問題データベースが必要不可欠であり、一定程度の基準で当該データベースの構築が必要である。

B 問題形式及び問題数

(1) 現状（PBT 方式）

解答方式には、多肢選択式の試験又は記述式の試験がある。問題数は試験の種類によって試験科目が異なっているが、概ね次のような範囲で出題している。

多肢選択式：大問 10～20 問

記述式：大問 5～6 問

※ 上記の「大問」とは、当該問題の下に枝問があることを意味する。例えば、多肢選択式では、一つの問題で 3～5 の選択肢があり、その選択肢の解答内容に応じた組み合わせによって正答を解答するようにしている。

(2) CBT 方式で実施する場合の論点（非 PBT 方式）

a) 実施方法（案）

多肢選択式の問題であれば、属人性を排除できおり、効率的に結果が求められる。この場合の出題にあっては、現状の大問の下枝問を小問の単位で出題するように分割し、一問一答形式に見直す。出題数は、現状の実問題数と変更のないように設定する。

例 大問 20 問（枝問題が各大問に 4 問）の場合

小問に分割すると、20 問×4 問＝60 問とする

ここで後述する D 問題評価の課題となる新問の事前評価として、本試験に 5 問程度を加えて出題し、対応する方法もある。この場合、当該新問は採点対象から除くことになる。

b) 課題

現状実施している記述式については、事実上、非 PBT では困難である。現状と同等レベルを維持するための試験方法について検討が必要である。

また、国家試験で出題した問題を採点対象から除くことは一般的でないため、新問を採点対象から除くような方法については再考が必要である。

C 出題問題の選定基準

(1) 現状 (PBT 方式)

資格試験では、一定の能力水準に達しているか確認するための問題を出題している。

(例 甲種化学免状に関する試験問題に、丙種化学特別免状に関する試験問題は出題しない。)

(2) CBT 方式で実施する場合の論点 (非 PBT 方式)

a) 実施方法 (案)

基本的には、現状の手法から逸脱することはせず、一定の能力水準に合わせた難易度の問題を出題する。

なお、受験者の知識水準 (能力) を測定するため、適切に問題を出題する CAT も CBT 方式であれば可能である。

b) 課題

CAT は簡単な問題から高難易度の問題まで幅広く準備することが必要

D 問題評価

(1) 現状 (PBT 方式)

試験実施後に問題の正答率を算出し、当該問題の妥当性や難易度を評価している。

(2) CBT 方式で実施する場合の論点 (非 PBT 方式)

a) 実施方法 (案)

手法 1 : 試験実施後に問題の正答率を算出し、その問題の相応しさ (難易度) を評価する。

手法 2 : 試験実施前に、プレ試験等により受講者毎の問題の正答率を算出し、その問題の相応しさ (難易度) を評価する。

b) 課題

手法 1 は、試験後の評価であるため、事後的な評価以上の活用が難しく、受講者ごとに問題の難易度のバラつきが大きくなる。

手法 2 は、試験前の評価であるため、プレ試験等により試行回数の蓄えを行うため、評価する人数の確保が必要である。

E 採点

(1) 現状 (PBT 方式)

上記 B で記したとおり、問題形式には多肢選択式と記述式があり、それぞれ採点方法はことなる。

多肢選択式は、あらかじめ定めた配点に従い、正答に応じて足し上げることで客観的に採点ができている。一方で、記述式は、採点時に偶然誤差が生じることから、対策を講じた上で実施している。

(2) CBT 方式で実施する場合の論点 (非 PBT 方式)

a) 実施方法 (案)

手法 1 : 正解した問題の配点を加算して採点する。

手法 2 : 等化处理し、受験者の能力の点数に換算して採点する。

b) 課題

手法 2 の場合、等化处理及び受験者の能力の点数化が必要である。

F 合否判定基準

(1) 現状 (PBT 方式)

問題形式にかかわらず、試験の試験科目ごとに一定以上の割合 (例 概ね 6 割) 以上の得点が得られれば合格としている。

(2) CBT 方式で実施する場合の論点 (非 PBT 方式)

a) 実施方法 (案)

手法 1 : 各科目 6 割以上の得点が得られれば合格とする。

手法 2 : 各問題の難易度を評価し、難易度における得点を配分の上、現状の合格基準と同等の得点が得られれば合格とする。

b) 課題

手法 2 の場合、「ある一定の能力レベル=合格」の基準を定めることが必要である。

なお、複雑な計算問題を解く記述問題は、途中点等の概念もあるため、手法 2 の基準を定めることは容易ではない。

G 試験時間

(1) 現状 (PBT 方式)

試験の種類や試験科目によって試験時間を定めている。60 分、90 分、120 分の 3 パターンである。

(2) CBT 方式で実施する場合の論点 (非 PBT 方式)

a) 実施方法 (案)

基本的な試験時間の変更はしない。各科目 (1 ~ 3 科目) を全て受けさせる上での手法は以下のとおり。

手法 1 : 各科目を全て 1 日で受験させる。

手法 2 : 定められた期間の中であれば、科目毎に分散して受験を可とする。

b) 課題

手法 1：試験時間が長くなるため、テストセンターの空き状況に依存する。

手法 2：1 科目目が不合格の場合の対応の検討を行う必要がある。

H 試験回数（受験機会）

(1) 現状（PBT 方式）

国家試験は年に 1 回実施し、検定試験は年に 1～4 回実施している。したがって、受験希望者が国家試験の申込みを失念等で行わなかったときは、1 年間、資格取得ができない。

(2) CBT 方式で実施する場合の論点（非 PBT 方式）

a) 実施方法（案）

実施回数を増加することは可能であり、その方法も特定の期日を増やすのではなく、実施期間の中から希望の試験日時を選択できるようにすることで、さらに受験機会を増やすことができる。

b) 課題

期間中における受験可能回数の制限の有無を検討する必要がある。また、十分な問題量の確保し、問題のデータベースの構築が必要不可欠である。

I 問題の公開

(1) 現状（PBT 方式）

高圧ガス保安法に基づく国家試験等は、高圧ガス保安協会のホームページで無償公開している。

(2) CBT 方式で実施する場合の論点（非 PBT 方式）

a) 実施方法（案）

受験者単位で試験問題が異なるため、試験問題の公開は困難である。対処としては、試験のレベルを推し量れるような参考問題の公開がある。

b) 課題

上記 A のとおり、問題データベースから出題する関係から、問題総量が十分でない時期での公開は、出題される問題が推定されるため、試験の妥当性に影響を与える可能性がある。

J 作問

(1) 現状（PBT 方式）

上記 A のとおり、同じ試験の同じ試験科目については、同一時刻に開始し同一時刻に終了するため、毎年、当該試験科目に係る問題を 1 種類について、外部専門家等による委員会での審議、作成し、出題している。

(2) CBT 方式で実施する場合の論点（非 PBT 方式）

a) 実施方法（案）

上記 D のとおり、問題の評価を行った後、問題データベースの構築及びメンテナンス

を行う。また、新問の作成は現状と同様の方法で実施する。

b) 課題

どの程度問題を準備するかによって、問題のデータベース構築及びメンテナンスの労力が変わるため、総労力が見えにくい。なお、新問についても、上記Dにあるとおり問題の評価が必要である。

K 受験会場

(1) 現状 (PBT 方式)

大規模会場にて、各試験を同一日時にて実施している。

(2) CBT 方式で実施する場合の論点 (非 PBT 方式)

a) 実施方法 (案)

多数の小規模会場 (テストセンター) にて、時間をずらして実施する。

b) 課題

地域的なハンディキャップによる経済的負担に配慮した会場設定が必要である。

次に、非 PBT 方式で国家試験等を実施する場合の制度的制約を確認した結果、高圧ガス保安法試験則第9条において規定されている『製造保安責任者試験及び販売主任者試験は、筆記による学科試験』の文言を改める必要があると考えた。そこで、既に非 PBT 試験を導入している情報処理技術者試験における情報処理の促進に関する法律施行規則を参考にし、以下のとおり改正案文を示した。

高圧ガス保安法に基づく高圧ガス製造保安責任者試験等に関する規則

改正前	改正後
(試験科目等) 第九条 製造保安責任者試験及び販売主任者試験は、筆記による学科試験とし、その試験科目は、次の表の製造保安責任者免状及び販売主任者免状の種類の欄に掲げる製造保安責任者免状及び販売主任者免状の種類に応じて、それぞれ同表の試験科目の欄に掲げるものとする。 2～4 略	(試験科目等) 第九条 製造保安責任者試験及び販売主任者試験は、筆記による学科試験又は電子計算機その他の機器を使用して行う学科試験とし、その試験科目は、次の表の製造保安責任者免状及び販売主任者免状の種類の欄に掲げる製造保安責任者免状及び販売主任者免状の種類に応じて、それぞれ同表の試験科目の欄に掲げるものとする。 2～4 略

②「国家試験等実施機関等の第三者による受験者本人の確認方法」のまとめ

最後に、国家試験等実施期間等の第三者による受験者本人の確認方法について、CBT 方式における本人確認は、会場受付で写真付の公的身分証明書と本人の2点を確認する又は事前提出さ

せた写真と写真無の公的身分証明書と本人の3点を確認する方法が、本人性の確認と同一性の確認を満たすことができる方法と考えられる。また、IBT試験においても同様に、事前に登録した写真付の公的身分証明書とカメラを用いて監視する2点確認又は事前に登録した写真無の公的身分証明書と事前に登録した写真とカメラを用いて監視する3点確認の方法が適切と考えられる。

今後の課題

非PBT方式にて国家試験等を実施する上では、試験問題を作成し、一定数の問題をプール及びその問題を評価するといった事前準備が必要不可欠である。事前準備には、モニター受験者によるパイロット試験があり、その結果を評価することで、従来のPBT方式での試験との連続性を確認する必要がある。

付録書 A

1. テスト理論概説

1. 1 テスト理論がなぜ必要か

多肢選択式や記述式、パフォーマンステストのそれぞれについて、テスト実施から採点、得点の算出までの流れを記すと、以下のようになる。

パフォーマンステスト：受験者からパフォーマンスのデータ（実技のデータなど）を収集する→評価者があらかじめ定められた評価基準に基づき、評定点をつける→評定点を集計し、受験者ごとに得点を算出する。

記述式テスト：受験者から記述式答案を集める→採点者があらかじめ定められた採点基準に基づき、採点する→採点結果を集計し、受験者ごとに得点を算出する。

多肢選択式テスト：受験者から解答データ（項目ごとに、どの選択肢を選んだかを記録したデータ）を収集する→解答データを正答キー（正答選択肢）と突き合わせ、採点する→必要に応じて配点をかけ、受験者ごとの得点を算出する。

パフォーマンステストと記述式テストは、評価者（採点者）が評価（採点）することをはじめ、その多くのプロセスが共通した要素を含むことがここからわかる。一方で、多肢選択式テストは評価者（採点者）がいらないため、より客観的な採点が可能であることから「客観式テスト」と呼ばれることがある。ただし、出題内容や選択肢の構成によっては「正解」「不正解」が明確に分かれない場合があるため、多肢選択式テストであっても出題内容の吟味を重ねる必要性は残る。

いずれのテスト形式とするかを問わず、テストは「人間の能力の一断面を測定するための尺度（ものさし）を構成するためのもの」ということができる。英語のテストは「人間の英語能力を測定するためのもの」であり、出題される問題項目は「人間の英語能力を測定するための道具」と位置付けられる。

ここで述べた3種の形式のテストのうちどの形式をとるにしても、採点結果（得点）が、受験者が実際にもっている能力を正確に反映した得点を返すことが求められている。評価者（採点者）が評価するテストの場合は、その評価者の質が一定水準に満たないと、受験者の能力を正確に反映した評価とはならなくなるおそれがある。また多肢選択式テストの場合は、項目の質が悪いと、受験者の能力とは関係ない要因で正解・不正解が分かれることとなってしまう、テストの質が悪化する原因となってしまう。

「テストの質」の良し悪しを評価するためには、出題された項目の内容的分析だけではなく、受験者の解答行動や評価者の評価をもとにした量的データに基づく分析が行われる。この量的データの分析を行う枠組みとして提案されたのが「テスト理論」である。

1. 2 古典的テスト理論

実際にテストを行うと、受験者ごとに得点が求められるが、得点は受験者ごとの出来の良し悪しを反映してばらついていることが期待される。理論的には、「受験者の能力のばらつき」に応じた「得点のばらつき」となるようなテストを用意できれば、もっとも良いであろう。しかし現実には、受験者の能力のばらつき以外の要因で得点がばらついてしまう場合も多い。

「テストの質」を評価する一手法として、テスト得点のばらつきが本当に受験者の能力、実力の

ばらつきを反映しているかを評価することが必要であり、そのための評価指標として以下の式

$$\rho = \frac{\text{観測された得点のばらつき}}{\text{真の得点のばらつき}}$$

に基づく「信頼性係数」 ρ を求める手法が提案されている。ここで「真の得点」とは、受験者がもっている真の実力を反映した得点であり、テストの結果からは観測されていないため、観測されたテスト得点から推定することとなる。この手法によるテストの質の評価は（次項に述べる）項目反応理論が提示されるよりも前から活用されていることから、「古典的テスト理論」に基づく評価と呼ばれている。

古典的テスト理論による評価は、テスト版（項目を束ねて受験者に提示する一連のテスト項目群、問題冊子）それぞれに対して行われる。たとえば、あるテストで10項目からなるテスト版を用いてテストを行った結果から、受験者の「真の得点のばらつき」を推定すれば、信頼性係数を推定することができる。一般に、互いに相関の高い項目（ある項目Aに対する正誤が、別の項目Bへの正誤と正の相関をもっている項目のペア）を集めることができれば、その項目群を提示した場合の信頼性係数は高まる。そのため、10項目よりも多くの項目を用いてテストを行えば、より質の高いテストが実現できる。ただし現実には試験時間を無限に取ることは不可能であり、受験者に提示できる項目の数には限りがあるため、項目間の正誤の相関を高める努力をすることで、高い信頼性を目指すアプローチをとることが現実的である。

1. 3 項目反応理論 (item response theory, IRT) ⁴²

古典的テスト理論による項目の質の議論においては、個別の項目一問一問に対する直接的な評価によるものではなく、あくまでテスト版全体に一つの「信頼性係数」が算出されるという評価がなされる。一方で、テストの実務的観点からは「項目ごとにその質を判断できる指標がほしい」という要求があった。

また、項目の困難度指標（問題項目の難しさの指標）として、古典的テスト理論では「正答率」（全受験者に占める、正答者数の割合。「通過率」ともいう）を用いる場合がほとんどであった。しかし、正答率を直接「項目ごとの困難度指標」とみなすうえでは、以下に述べる問題点が指摘されている。

たとえば、中学3年生に対して、学習指導要領上で中学3年生レベルとされる内容の項目を10問提示した場合、正解が多かった生徒は「中3レベルの学力をもつ」と判断されるであろう。一方で、小学6年生レベルとされる内容の項目を10問提示して、正解が多かった受験者は「小6レベルの学力をもつ」と判断できる。しかし、仮に出題された項目の難易度レベルが不明であるなら、正答の数を数えたとしても、その生徒の学力レベルがどの程度なのかを見極めることはできない。

一方で、ある10問からなるテスト版を提示し、正答数が「10問」であったとしても、その意味は「どの学力レベルの受験者が解答したか」によって変わることとなろう。すなわち、ある項

⁴² 項目反応理論全般を扱った書籍として、下記のものがある。

加藤健太郎・山田剛史・川端一光（2014）. Rによる項目反応理論 オーム社

目が「小6に対して正答率が高かった」のか、「中3に対して正答率が高かった」のかによって、困難度は変わってくる。しかし、実際に観測できるのは「多数の正解が得られた」という点で同一であり、やはりここでも正答率を算出しただけでは、学年をまたいだ「標準化された困難度」を求めることができないことがわかる。

難しい問題ばかり 10 問を用意してテストをした場合と、やさしい問題 10 問でテストをした場合で、受験者の正答数だけからでは、その受験者の能力を相互に比較できないことがわかる。しかし、これら 2 種類のテスト版で共通の「能力尺度」を潜在的に仮定し、この尺度の上で受験者の能力を表せば、この能力尺度上で 2 種類のテスト版に対する正答数を位置付けることが可能となる。

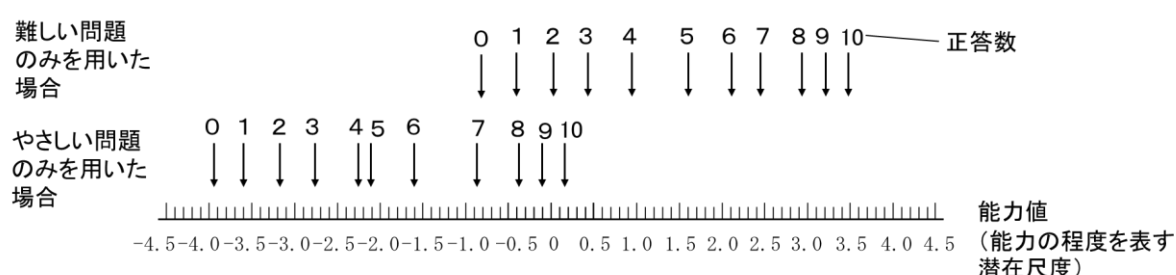


図1 2種類のテスト版(難しい問題10問とやさしい問題10問)への正答数と、潜在特性値(能力値)との対応関係を示した図

図1の「能力尺度」は、出題した項目の困難度に依存しないで定義され、潜在的にその存在が仮定される尺度であり、多数の項目に対する正誤が分かれた要因として、潜在的な一次元の能力尺度の大小が挙げられる可能性が高い場合に、そのような尺度の存在を考えることができる。言い換えれば、そのような一次元の「能力」＝「単一の学力」ではない要因で正誤がばらつくような正誤データに対しては、図1のような潜在尺度の存在を仮定できないこととなる。一次元の能力尺度を仮定することの是非は、心理学における「因子分析」を正誤データに対して適用することとて、検証することができる⁴³。

項目反応理論(IRT)においては、受験者が項目に正答する確率を、潜在的に仮定した能力尺度の上で表現する。具体的には図1のように、潜在的に仮定された「能力値」の尺度を横軸にとり、対応する期待正答確率を縦軸にとる。このような「項目反応関数」(または「項目特性曲線」)を出題した各項目について一つずつ考え、受験者の正誤データから、その形を推定する。

⁴³ 因子分析に関する技術的な詳細については下記の文献を参考のこと。

柳井晴夫・前川眞一・繁桝算男・市川雅教(1990). 因子分析 ―その理論と方法 朝倉書店

また、IRTが1因子を仮定した因子分析の一種であることや、因子分析により一因子性を検討する事例については、以下の文献を参照のこと。

光永悠彦[編著] 西田亜希子[著](2022). テストは何のためにあるのか ―項目反応理論から入試制度を考える, pp. 71-76.

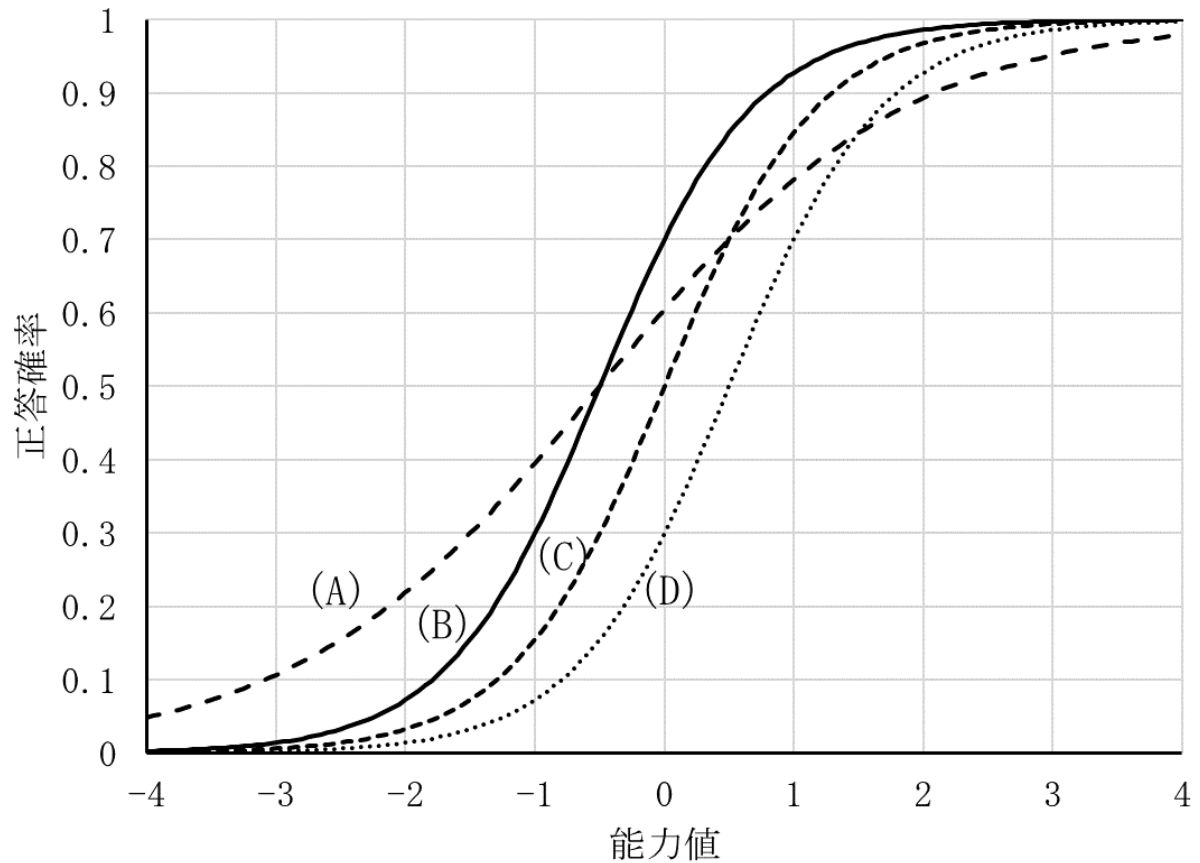


図2 項目反応関数の例

この図は4項目について示した。(A)は $(a, b) = (0.5, -0.5)$ 、(B)は $(a, b) = (1.0, -0.5)$ 、(C)は $(a, b) = (1.0, 0.0)$ 、(D)は $(a, b) = (1.0, 0.5)$ の場合を表す。

もし実際に観測された正誤について、その背後に一次元の「能力」を仮定し、潜在的に仮定された能力の違いにより正誤が分かれていると判断できれば、IRTの枠組みを用いて、項目ごとに「困難度」を推定することが可能となる⁴⁴。

項目反応関数は「ロジスティック曲線」を用いることが多い。具体的には、以下の式が用いられる(2パラメタ・ロジスティックモデル)。

$$P_j(\theta) = \frac{1}{1 + \exp(-Da_j(\theta - b_j))}$$

⁴⁴ また、項目ごとに「識別力」と呼ばれる指標を推定することも可能である。能力値が近い2人の受験者AとBにおいて、Aの正答確率よりもBの正答確率の方が顕著に大きくなった場合、その項目はAとBの能力をより弁別できると考えることができ、「識別力」が大きな項目であるとみなされる。識別力は、項目反応関数の変曲点における接線の傾きに該当する。「識別力」はすべての項目において同一の値(または「1」に固定)とするという場合もあり、「1パラメタ・ロジスティックモデル」と呼ぶ。1パラメタと2パラメタのいずれのモデルを用いるかについては理論的・実践的観点からの議論があるが、詳細は紙幅の都合で割愛する。

ただし $P_j(\theta)$ は項目 j に正答する期待確率、 a_j は項目 j の識別力、 b_j は項目 j の困難度、 θ は受験者の能力値、 D はロジスティック曲線を正規累積曲線に近似させるための尺度定数を表す⁴⁵。

項目反応理論による分析によって、受験者の能力分布によらない、標準化された困難度（や識別力など）の指標を推定することが可能となる。また、解答した項目の困難度のバリエーションによらない受験者の能力値を推定できる。さらに、評価者の情報を加味した項目反応理論による分析の枠組みも提案されている（Many-Facet Rasch Model, MFRM）。

項目の困難度、受験者の能力値は、図3に示すように、一次元の潜在変数としての「能力値」の軸上で表すことができる。このことは、困難度と能力値が互いに対応していることを表している⁴⁶が、MFRM による評価者の「評定の厳しさ」指標も、これと同列に扱うことができる。

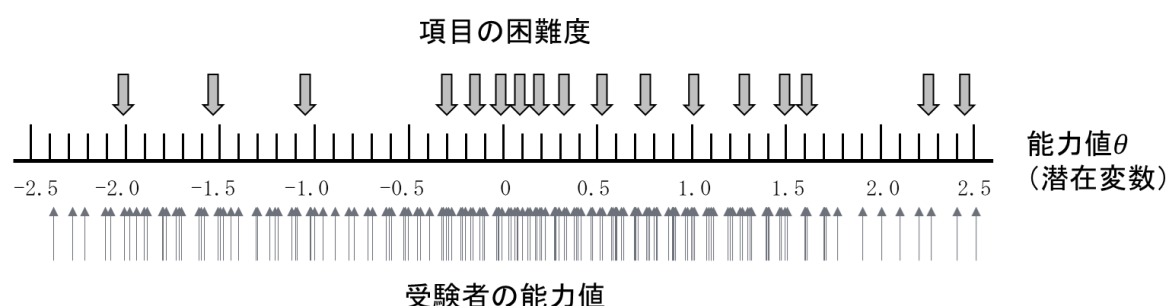


図3 IRT における、困難度と能力値の関係

ただし、IRT による分析結果が、どのような正誤データに対しても妥当な結果となるわけではない。IRT の結果が理論通りの意味を持つための「前提」があることが知られている。もっとも重要な前提としては、前述の、能力が潜在的な一次元の尺度を構成しているというものがあるが、これと実質的に同じ意味を持つ前提として「局所独立の仮定」がある。これは同じ能力をもつ受験者の間で、出題された項目群への正誤は相互に独立して起こるという仮定である。たとえばあるテストにおいて、1問目で導いた正解を用いて2問目を解くという構造の項目を出題した場合、2問目に対して正解した受験者は、1問目に対しても正解しているということとなり、能力の違いではない要因（2問目を解く手がかりが1問目の正解であるという要因）に依存して2問目の正誤が決まってしまう。IRT の分析にかけるためには、このような局所依存関係にある項目群がないように、小問を集合させた形式の項目をテスト版に集積させることが多い。

⁴⁵ 困難度や識別力のことを、項目ごとの特性を表す値ということで「項目特性値」や「項目パラメタ」と呼ぶ。項目特性値の推定手法については、以下を参照のこと。

豊田秀樹（2012）. 項目反応理論[入門編]（第2版） 朝倉書店, pp.65-84.

⁴⁶ 項目の難易度を正答率（通過率）で表し、受験者の能力を当該受験者の正答率で表した場合、前者の分母は「全解答者数」であり、後者の分母は「全解答項目数」であるため、これらの「正答率」は実質的に異なる意味をもっている。したがって、正答数（や正答率）を求めた場合、同じ軸上でその値を表現することはできない。

1. 4 複数のテスト尺度の共通尺度化と予備テストの役割

年に複数回行われるテストの結果は、IRT による能力値を用いると、出題した項目の困難度に依存しない形で得点を表示できる。しかし、ある実施回（第1回）での IRT による分析で見出される能力尺度は、その回の分析の中でのみ有効である。他の実施回（第2回）において、別の IRT による分析を行って見いだされた尺度とは、直接比較することができない。

ただし、第1回テストと第2回テストの間で、共通の項目を含ませるか、又は共通の受験者が存在する場合、それらの共通要素を手がかりとして、第1回テストと第2回テストの間で共通な能力尺度を構成することが可能となる。すなわち、第1回テストの能力値と第2回テストの能力値が比較可能となる。

異なるテスト得点尺度を、共通の尺度上で表現する操作を「等化」(equating)と呼ぶ。等化を行うためには、異なるテストがどれも同じ概念を測っており、妥当性が高いことが要求されている。測定されている概念の同一性が保たれていない場合における共通尺度化を「リンキング」と呼んで区別することもある。

通常、一度実施したテストの得点を、他のテスト実施後に修正することは、受験者に対して混乱を招くことから、行われない。そのため、第1回テスト実施後に行われるすべてのテストについて、その能力尺度を第1回テストの能力尺度に変換する。第2回テストが実施され、IRT による分析により能力尺度が構成されたら、その尺度を第1回テストの尺度上に変換し、変換後の尺度上で第2回テストの得点を表示する。第1回テストの能力尺度は固定して動かさないため、この操作により、すべての回次のテストが第1回テストの能力尺度上で表されることとなり、異なる回のテスト得点が同じ尺度上で表現することができる（図4）。

実際にテストを行う場面では、能力の基準となる第1回テストは、受験者に成績を返さず、基準を設定する目的や、テスト項目の妥当性や信頼性を検証し、テストの仕組みが当初の期待通りに機能するかを検証する目的で行われる。受験者に得点を返す目的ではないテストを「予備テスト」や「フィールドテスト」、「プレテスト」と呼び、予備テストの受験者は「モニター受験者」と呼ばれる。

予備テストによって構成された能力尺度は、その後に行われるテストで、能力尺度の基準となる役割を果たす。そのような基準となる受験者集団を「規準集団」と呼ぶ。等化の操作により、毎回のテスト得点尺度を規準集団上の尺度で表現し、互いに比較可能な得点として扱うことができるようになる。ただしどのような場合においても等化が可能というわけではなく、どのテストで測定される概念も同じであり、尺度の信頼性や妥当性が高いことなどの「等化の前提」があること⁴⁷が知られている。

⁴⁷ 等化の前提については、下記の文献を参照のこと。

川端一光（2014）. 等化 加藤健太郎・山田剛史・川端一光[著] Rによる項目反応理論 オーム社, pp.257-258.

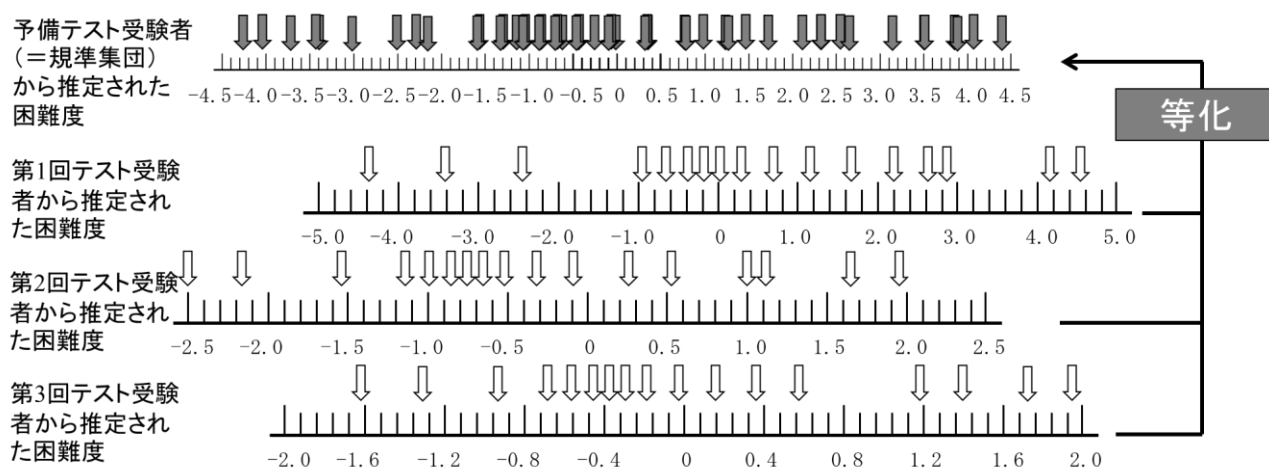


図4 予備テスト尺度による複数回テストの能力値の共通尺度化（等化）

2. 知識・技能の測定方法

2. 1 知識・技能の構成要素を明らかにする（テストの妥当化）

テストの結果が社会的に活用されるためには、そのテストが「テストが測りたいもの」を的確に測定していることが求められよう。たとえば、英語のテストとして設計されたテストが、受験者の英語能力ではなく、注意力を測定したテストであるならば、英語能力の高い受験者においても注意力の欠如により「うっかりミス」が発生して不正解が続出し、英語能力とは関係ない要素で正誤がばらつくことになってしまう。

先に「テストは尺度を構成するもの」と述べたが、「テストが測りたいもの」を的確に測っている尺度となっていれば、その尺度は「妥当性が高い」ということができる。測りたい要素を的確に測っているテスト版を用いてテストを行えば、妥当性の高いテスト尺度となることが期待できる。

テストの妥当性を決める要素としては、「構成概念妥当性」の向上を目指す考え方が主流となっている。たとえば、テストで測りたいものを一つに絞ったにもかかわらず、実際に測定されている概念が複数ある場合、正誤データを因子分析すると、能力尺度として次元ではない可能性が指摘される。このことは、尺度の妥当性を低下させる要因となる。また、英語テストXを開発する過程で、受験者にテストXとTOEICを同時期に受験させ、その得点がほとんど無相関であった場合、TOEICはすでに「英語のテスト」として定評があることから、英語テストXが本当に英語能力を反映した得点を返しているか疑問であるということとなり、妥当性が低い、と判断される。

これまで日本においては、テスト項目を作成後、直ちに受験者に提示し、解答データを収集してきた。しかし多くの妥当性検証は、実際の受験者から解答データを収集したうえで、データに基づき行われる。もちろん、項目内容を質的に判断し、項目ごとに問うている内容の評価を行うことも妥当性検証の一部に含まれるが、それは妥当性検証の一つに過ぎない。これらの妥当性検証の過程を通じて、テスト尺度の妥当性を向上させることを「テストの妥当化」と呼ぶ。

2. 2 より確からしい結果となるテストの仕組み（テストの信頼性向上）

テストの「質」を左右する要素としては、前述の「信頼性係数」や「識別力」といった指標による検証も不可欠である。テストの「信頼性」とは、テスト版を能力の測定の道具と考えたとき、その測定道具の性能を示していると解釈される。

たとえば、あるテストXで500点をとることができる受験者がいたとして、この受験者がテストXの異なる種類のテスト版を10回受験し、毎回において499点から501点の範囲に得点がばらついたらとすれば、これら10種類の測定はきわめて信頼性が高いといえる。しかし現実にはより広い範囲に得点がばらつくことが予想される。その範囲がプラスマイナス50点なのか、100点なのかによっても、テストの「質」を論じることが可能であろう。

テストの信頼性を向上させるためには、前述の通り、項目間で高い正誤の相関となるテスト版を用いることが必要である。このことは、テスト項目の内容の吟味だけではなく、テストで測りたい概念の範囲をある程度絞ることも要求している。一つのテストで一度に測れる範囲は、素朴に考えるほど広いわけではなく、ごく狭い範囲に限られる。一方で、幅広い概念をまんべんなく測定するためには、それだけ多くの項目が必要となる。信頼性を高めようとする測定範囲が狭くならざるを得ず、ある程度のあいまい性をもった広い概念を測ろうとすれば、信頼性が低くなる。これは「帯域忠実度ジレンマ」と呼ばれている。

尺度の信頼性と類似する概念に「テスト情報量」というものがある。テスト情報量とは、受験者に出題したテスト版が、受験者の能力をどの程度弁別できるかを示しており、能力値の関数として与えられる。IRTを用いた分析を行うと、能力レベル別に、出題したテスト版の情報量がグラフの形で描くことができる。図5に示すグラフを「テスト情報量曲線」と呼ぶ。

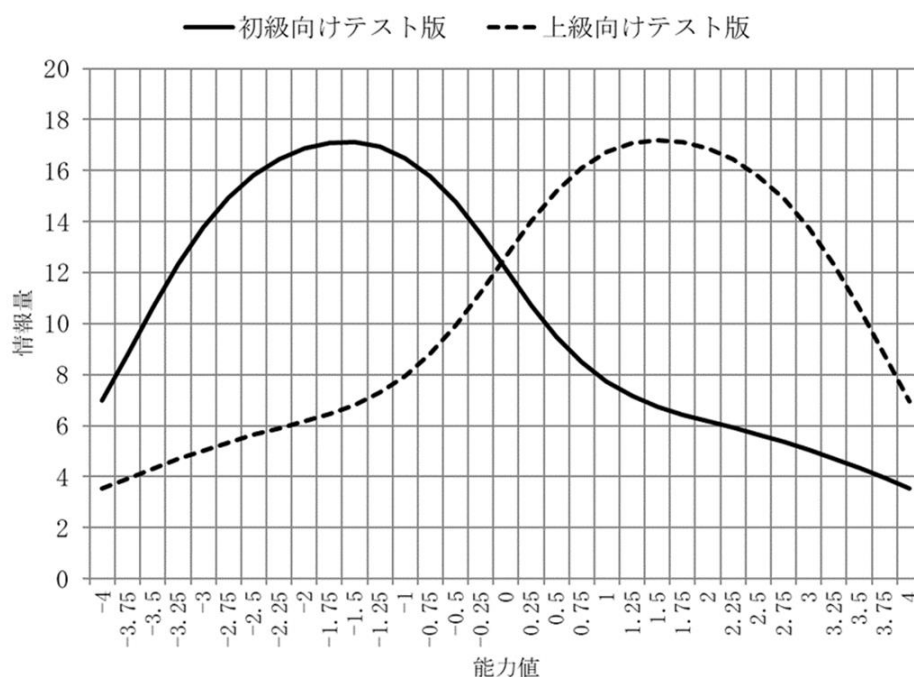


図5 テスト情報量曲線の例

困難度の分布が異なる2種類のテスト版について示した。

テスト情報量曲線は、出題したテスト版ごとに IRT 分析を行い、困難度や識別力を推定することとて、描くことができる。たとえば、困難度が低めの項目から構成されるテスト版を IRT 分析すると、図 5 の実線に示す曲線となり、困難度が高めの項目の場合は図 5 の点線のような形を呈する。すなわち、困難度が低めのテスト版は、能力値が低い受験者に対して、その能力値を精度よく推定することが可能となり、困難度が高めのテスト版は、能力値が高い受験者に対して精度が高く能力値を推定することができる。

テスト情報量の考え方を導入することにより、コンピュータ適応型テストを実現することができるようになる。また適応型テストではない場合であっても、出題されたテスト版が初級者向けだったのか、それとも上級者向けだったのかについて、大まかな傾向を検討する材料となる。

2. 3 より役に立つテストを設計する（テストの有用性向上）

テストが年 1 回、東京のみで行われるよりも、年複数回行われ、全国各地で受験可能であるならば、受験者も受験しやすくなるであろう。実用に供されるテストにおいては、受験者からみて有用であるテストと、そうではないテストに分けられる。このようなテストの違いを「有用性の違い」と呼ぶ。

有用性の程度は、テストを取り巻く立場の違いによっても変わりうる。あるテストを年複数回実施することとすれば、受験者にとって有用ではあるが、テスト実施機関からすれば会場の確保や試験監督者の配置といったコストが余計にかかることとなり、有用性は損なわれる。万人にとって有用なテストというものは存在せず、一方の立場からの有用性が高まれば、時として一方の立場に過大な負担をかけることにつながりうる。

一方で、コンピュータを用いたテスト（Computer Based Testing, CBT）を導入することで、テスト実施のために問題冊子を印刷する手間がなくなり、より現実的なテスト制度を構築できる。新しい技術の導入や実施上の工夫により、多くの立場の者にとって有用性の高いテストを実施できる可能性がある。有用性の評価については、受験者に対してアンケート調査を行うことや、外部の有識者による検討、作問者に対する問題項目の作問の難しさに関するインタビュー調査といった手法により行われる。

2. 4 資格試験におけるテストの「質」とテスト形式

これまで、テスト一般におけるテストの「質」について議論してきたが、資格試験の場合において「質の高いテスト」は、以下のような特徴を備えたテストであるといえる。

- 高い公平性が保たれたテスト。
- 妥当性及び信頼性が高いテスト。
- 合格・不合格だけでなく、学習の過程でどの位置にいるかを、受験者が知ることができるテスト。

資格試験においては、資格を認定する公的機関がテストを実施するという関係上、公平性が最

優先されることが多い。また、受験者の本人確認や、受験資格の確認といった要素が重視される傾向にある。CBT化を実現する上では、IRTを導入するか否かに関わらず、オンラインによる本人確認手段・試験中の解答行動監視手段の確立が、一つのハードルとなることが予想される。

上記のような特徴をすべて備えたテストを構築することは、実施のためのコストや人的資源の制約、受験者数の制約などを考えると、現実には難しい。また、CBTとすることにより公平性が損なわれるおそれがある場合、紙ベースのテスト(PBT)のままで実施するという選択もありうる。構築すべきテスト制度について、何が優先されるかを十分検討したうえで、テスト制度構築に臨むべきである。

多くの資格試験においては、テスト形式として「多肢選択式」と「記述式」が用いられている。それぞれに一長一短があることが知られている⁴⁸。

【多肢選択式】

(利点)

- 一度に実施できる項目数を多くでき、項目数が少ないための欠点を補うことができる。
- さまざまな知識や理解を一度に問うことができるため、出題内容のバランスをとりやすく、公平性も確保しやすい。
- 採点作業が客観的に行える。採点作業の効率もよい。
- 選択肢の工夫で、知識にとどまらず高度な思考やそのプロセスも問える。

(欠点)

- 選択肢に盛り込まれなかった事項に関しては何も情報が得られない。
- 選択肢を示すことで、誤答に気づいたり正解を見て思い出したりする受験者が出てくる。

【記述式】(解答を単語で答えさせる短答式ではなく、比較的長い文章を書かせる解答形式)

(利点)

- 情報を組み合わせたり、何段階かを経て結論を導いたり、絶対的な正解がないような内容、とくに日常的・現実的な場面を設定して知識を応用する力を見たり、自分の考えをまとめて表現させたりするのに適している。

(欠点)

- 1回のテストで課すことのできる項目数が限られる。
- 素材の選び方、公平性、出題のバランスには多肢選択式以上に気をつける。
- 採点の主観性や一貫性、一致性への対策が必要になる。採点作業に多大な労力と時間がかかる。

上記の利点と欠点を考慮し、多肢選択式と記述式のテストを組み合わせることで、より質の高いテストを構成することが可能となる。ここで「質の高いテスト」とは、さまざまな能力レベル

⁴⁸ 以下の利点と欠点の記述は、下記の資料によった。

日本テスト学会[編] (2010). 見直そう、テストを支える基本の技術と教育 金子書房, p.28.

をもつ受験者が混在している受験者群に対して、比較的多数の評価のポイントについて公平性をもった能力の判定基準を与えるようなテストを指す。

まず、受験者全員に対して、客観的テスト（多肢選択式または短答式）を用いて、基礎的な事項に関する知識や、それらの知識を組み合わせで適切に判断する能力を「第一次試験」として測る。その結果、十分な基礎能力を有すると判断された受験者だけを選抜する。その後「第二次試験」として、記述式のテストにより、より高度で現実場面に即した具体的な実務能力を測定する。

このテストの方法により、第一次試験の受験者が多数であり、それら全員に対して第二次試験で問う内容をテストできないという制約をクリアすることができる。また、第二次試験において、限られた受験者に対してのみ記述式の採点を行えばよいということで、多大な労力がかかる記述式答案の採点作業を少数の採点者で行えるという利点がある。また、第一次試験で比較幅広い分野にわたるさまざまな素材を用いてテストを行うことができ、テストで考慮すべき評価ポイントをなるべく幅広くカバーすることが可能となる。

2. 5 パフォーマンスを測るための方法と尺度化

人間のふるまいや行動等に対して、何らかの評価が下され、その結果をもとに成績が判断されるテストを「パフォーマンステスト」と呼ぶ。パフォーマンステストでは図6に示すように、得点を算出する上で採点基準を定める必要があり、またこの採点基準にしたがって人間のふるまいや行動等を評価する「評価者」により、公平な形で評価が下される必要がある。これらの点は、客観式テストにはない特徴である。

パフォーマンステストによって、問題項目への正誤からではわからない、受験者の特性や能力を測定する可能性が開ける。たとえば英語による会話をする場面で、相手とどれだけ流暢に意思疎通を図ることができるか、といったテストによる測定が可能となる。ただし、測定される概念と測定方法、採点基準（評価基準、ルーブリックとも呼ばれる）を厳密に決め、テストで測られる能力尺度の妥当性を検証する必要がある。また、採点基準は、評価者が評価を下すための重要な手掛かりの一つとなるものであり、採点をする上での基礎資料となるとともに、「テスト対策」を防ぎ、公平なテストを実施するために、受験者にはその詳細を伏せることが多い。

客観式テストにおいては、出題される項目への正誤が、測りたい能力尺度の大小によって分かれるかどうかを検討するのに対し、パフォーマンステストにおいては、出題される項目だけではなく評価者の評価においても、その妥当性が問われる。その意味で、パフォーマンステストの妥当化は困難を伴う。評価者の評価の妥当性としては、評価が受験者ごとに一貫しているか、評価のポイントが採点基準の通りであるか、評価者に対して厳しすぎたり優しすぎたりしないか、といった観点から検討される。テスト実施に先立ち、模擬的な答案を評価させ、期待通りの評価が下せるようになるように、評価者に対して「評価者トレーニング」を行うことが多い。

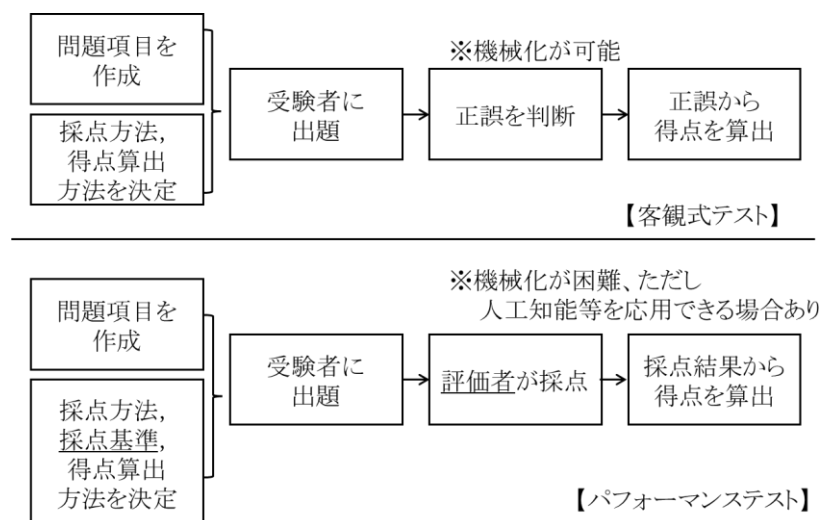


図6 客観式テストとパフォーマンステストの実施過程の違い⁴⁹

2. 6 基準設定 (standard setting) の方法と標準化テストへの適用

資格試験は通常、一度きりのテストだけで終わることはなく、複数回行われる。資格試験においては、ある一定の能力レベルを超えた受験者全員に「合格」という結果を通知するが、公平なテストとするためには、その「一定の能力レベル」のカットオフポイントを決定する方法が問題となる。

公平な資格試験を実施するうえでは、複数回行われる資格試験をまたいで、「ある一定の能力レベル＝合否基準」が同一であることが求められる。そのためには、まず規準集団となるモニター受験者に対して予備テストを行い、規準集団上で合否を分けるカットオフポイントを定めたうえで、複数回のテストの能力尺度を共通尺度化（等化）する（図7）。この手法により、カットオフポイントの決定を毎回の本試験を行うたびにを行う必要がなくなり、効率的なテストの実施につながる。また、カットオフポイントの設定について、本試験の実施とは別にテスト実施者が検討することができ、より公平なテストの実施が期待できる。

⁴⁹ 出所は以下の通り。ただし、一部改変している。

光永悠彦[編著] 西田亜希子[著] (2022). テストは何のためにあるのか ―項目反応理論から入試制度を考える― ナカニシヤ出版, p. 136

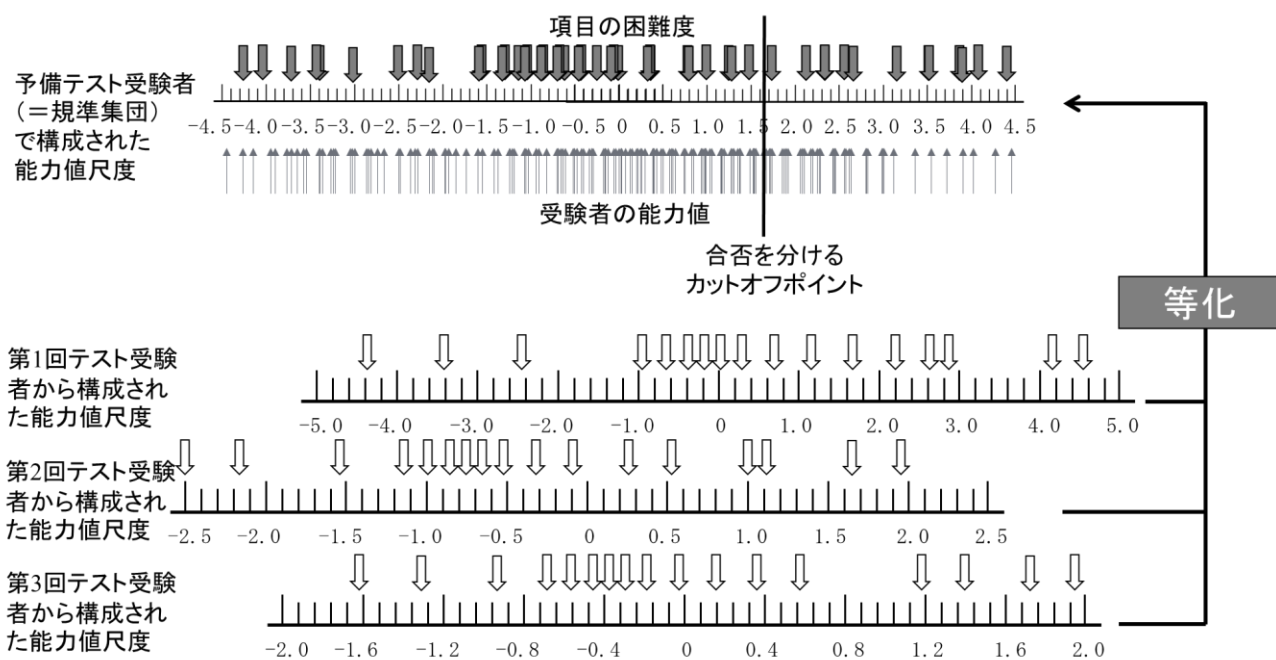


図7 複数のテストのカットオフポイントを等化によりそろえる方法

予備テストにおけるカットオフポイントの設定方法は、以下の4通りが知られている⁵⁰が、「問題文を複数の専門家が読んで判断する」過程を伴う2通りの方法（Angoff 法、Bookmark 法）と、専門家を要しない2通りの方法がある。

【Angoff 法】

複数の専門家に問題項目を1問ずつ提示し、「もしその項目に100人の平均的な能力を持つ受験者が解答したら、何人が正解できるか」を問う。その数値を項目ごと・専門家ごとに並べ、全体の平均をとる。ただし、同じ問いを2度以上繰り返すことで、専門家に再考の機会を与え、より確からしい回答を得ようとする場合が多い。項目ごと・専門家ごとの回答結果の平均値が「全体のテストを100人が受験した場合の、期待される正答数の平均」であり、「期待される正答率」と解釈できる。この値をもって、カットオフポイントとする。

たとえば10項目に10名の専門家が回答した結果、平均値が60.0となったとすると、これら10名の専門家は平均的な能力の持ち主であれば60%の正答が得られると期待している。そのため、10項目中6問正解以上を「合格」とすればよい。

【Bookmark 法】

IRTによる困難度が推定できた場合に適用可能な手法である。困難度が低いほうから高いほう

⁵⁰ 概要は以下を参照のこと。

Cizek, G.J. (小泉利恵訳) (2008). 規準の設定 Downing, S.M. & Haladyna, T.M. [編] 池田央[監訳] テスト作成ハンドブック 教育測定研究所, pp.244-278.

へ、順番に項目を並べたテスト版を用意し、複数の専門家に1問ずつ順に見せていく。「提示されていく項目が徐々に難しくなっていくので、これらの項目に100人が解答したとして、最初に1/3の受験者（33人）が不正解するだろうと考える項目に印をつけてほしい」と教示する。

たとえば、ある専門家が印をつけた項目の困難度が0.5で、識別力が1であったとする。そのような項目の項目特性関数は、図2の(D)に該当する。この図の縦軸が0.67（すなわち、100人中33名が不正解＝67名が正解）である場合に必要な能力値は、およそ0.9前後と求まる（実際には、項目特性関数の式に識別力と困難度を代入すれば、正確な値が求まる）。これらの能力値を複数の専門家ごとに求め、それらの値を平均した値をカットオフポイントとする。

また、専門家を要しない基準設定の手法として、下記の2種類が提案されている。いずれも専門家を要しない代わりに、特別な受験者集団を要する手法である。

【対比グループ法 (Contrasting Groups Method)】

ある資格をもつにふさわしい能力を持っていることが確実な受験者集団（習得者集団）と、資格取得の途上にある受験者集団（未収得者集団）をモニター受験者として、予備テストを行い、得点分布を度数分布表の形で集団別に表す。二つの集団を併合した度数分布は通常ふた山になるが、その山の谷間にあたる得点をカットオフポイントとする。

【境界グループ法 (Borderline Group Method)】

資格を持つにふさわしいかどうかの議論が分かれる受験者集団を考え、その受験者集団がテストを受験した場合の得点分布から、カットオフポイントを決める方法。得点分布の中央値がカットオフポイントとされる場合が多い。

テストによっては、複数の科目を含むものがある。次章で述べる資格試験はその一例で、複数の科目（次章の例では「サブ分野」）のそれぞれについて得点を合算するテストについて記述したが、複数の科目や分野を含むテストにおける合否判定方法としては、以下の2通りがある。

- 科目ごとの得点を合算して「総合点」を求め、総合点に対して設定されたカットオフポイントを上回っているか否かで合否を決定する。
- 科目ごとそれぞれにカットオフポイントを設け、カットオフポイントをすべての科目で上回った受験者を「合格」とする。

前者は科目全体の能力の大小を考慮して合否を決定する方法であり、科目ごとの能力に関係なく合否が決定され、カットオフポイントの設定も1つで済む。後者は科目ごとに必要とされる能力の尺度を、合否に反映させることができる方式であり、科目ごとにカットオフポイントの設定が必要である。

資格試験においては、テストで測られる能力の尺度が「科目」等の複数の単位に分かれており、資格試験の制度趣旨から、科目ごとに能力が基準を超えているかを判断する場合が多い。そのため、後者の手法により合否が判断される場合が一般的である。試験制度によっては、科目単位で

合否を判断し、すべての科目で合格とされて初めて資格が得られるという仕組みを取り入れることで、受験者に対する有用性が高まるかもしれない。

2. 7 効率の良いテストの実施手法：コンピュータ適応型テスト (CAT)

質の高いテストを実施するためには、高い信頼性をもつテスト版を用いてテストを行うことが求められる。そのためには、2. 2で述べた「テスト情報量曲線」の形が、どの能力値に対しても高い情報量となるようなテスト版を用いる必要がある。そのようなテスト版を構成するためには、たとえば困難度が -2 の項目を10問、 -1 の項目を10問、 0 の項目を10問、 1 の項目を10問、 2 の項目を10問というように、さまざまな困難度の項目を多く出題することが求められる。

しかしながら、前述のようなテスト版を用いた測定を行うと、たとえば能力値が -2 の受験者に対しては、困難度が 2 の項目を出題しても、「能力の低い受験者に対する難しい項目」を出題することになってしまい、不正解となることが分かっているにもかかわらずテストの仕組み上出題せざるを得ないこととなり、無駄が生じる。

一方で、視力検査の測定場面を考えてみると、眼鏡をかけている受験者に対しては 1.0 以上の項目を提示せず、低い視力の項目を多く提示して回答させることで、より効率的に検査を行っている。この考え方をテストにも導入すると、事前に受験者の能力レベルの見当をつけておき、能力値の初期値を求めたうえで、その能力値に近い困難度の項目を出題し、正誤反応の結果を見て次に出題する項目を決定するという測定方法を考えることができる。これが「コンピュータ適応型テスト」(Computer Adaptive Testing, CAT)の考え方にほかならない。

CATの仕組みを理解するためには、出題した項目の困難度と識別力、それに項目ごとの正誤状況を与えることで受験者の能力値が推定できることと、能力値の推定の際に、推定された能力値の確からしさ(能力値の標準誤差、95%信頼区間)も同時に推定でき、その値は通常、出題項目数が増えるごとに小さく(狭く)なっていくことを理解しておく必要がある。一般的なCATの仕組みは以下のとおりであるが、複数の分野が混在したテストの場合においては、必ずしもこの手順通りではない場合もある。

- ① 最初に、困難度や識別力が分かっている項目を数問(5問程度)出題し、能力値の推定値の初期値を求めておく。これが暫定的な受験者の能力レベルにあたる。
- ② 能力値の推定値の初期値に近い困難度の項目を出題し、正誤を判断する。もし正解していれば、より高めの困難度をもつ項目を次の出題候補とする。不正解の場合は、より低めの困難度の項目を次の出題候補とする。
- ③ の能力値推定において、能力値の推定値の標準誤差も推定されるため、この値がある閾値を下回っているかどうかを判定する。もし閾値を下回っていれば、これ以上の出題をしても能力値推定値が動かないと判断し、テストを打ち切り、その時点での能力値推定値を受験者の得点算出の根拠とする。閾値以上であれば、②に戻って出題を続ける。

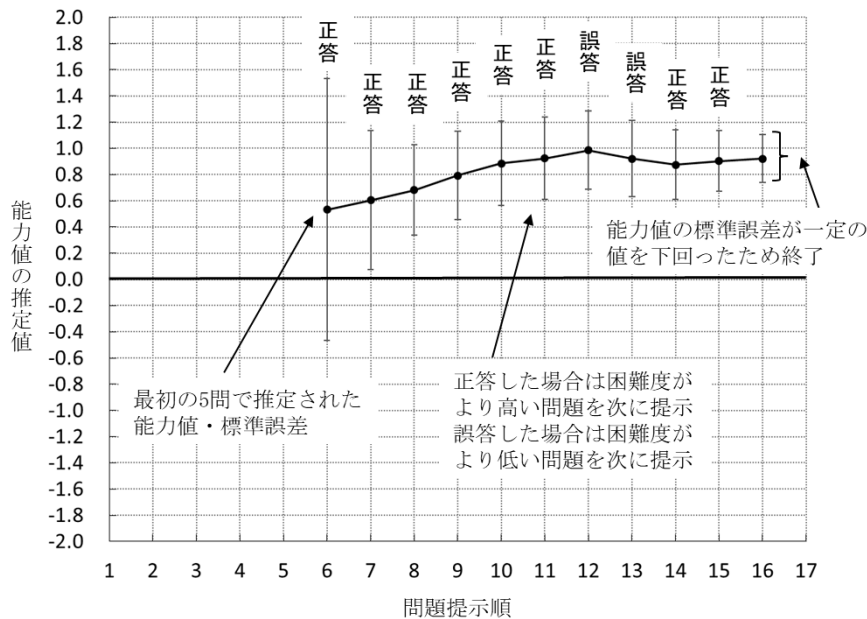


図8 CATの実施プロセスに伴う受験者の能力値推定値の推移⁵¹

図8は、CATにより能力値を推定するプロセスにより、受験者の能力値の推定値がどのように更新されていくかを表したものである。出題される項目の識別力が高いと、標準誤差の減少幅が大きくなるが、一般に高い識別力の項目は限られる（性能の良い項目は数に限りがある）ため、出題される項目の多様性を重視して、識別力が低い項目をあえて出題するアルゴリズムを採用することもありうる。

また、次に出題する項目を選抜するアルゴリズムには数種が提案されているが、その多くは「次に出題することで、これまで出題した項目のテスト版から求めたテスト情報量が最大となる項目」というように、テスト情報量を基準としたものである。CATを実施するためには、さまざまな困難度や識別力をもつ項目群（最低でも数百問以上）を「項目バンク」のなかに用意しなければならないが、その項目バンク全体から求めたテスト情報量曲線が、どの能力値に対しても十分に高いことが理想である。

どの受験者に対しても同一のテスト版を用いてテストを実施する場合、受験者の幅広いさまざまな能力値に対応して、さまざまな困難度の値をもつ項目を出題することで、どの受験者に対しても情報量が高い（すなわち、精度の良い、標準誤差の小さい）測定が期待できる。しかし、そのためには多くの項目を出題しなければならない。CATの仕組みによって、これまでさまざまな困難度の値の項目をまんべんなく全員に出題していたテストを、受験者の能力に応じて個別最適化された項目によるテストとすることがができる。これにより、能力の判断に必要な項目数を削減しつつ、非CATのテストと同程度またはそれ以上の能力値の推定精度とすることが期待される。

⁵¹ 出所は以下のとおり。

光永悠彦（2017）．テストは何を測るのか—項目反応理論の考え方 ナカニシヤ出版，p. 187

3. CBT 方式の試験に関する詳細（理論に基づく想定例）

3. 1 試験実施の架空例：項目バンクの構築と予備テストの実施

本項以下、架空の資格試験を題材に、より質の高い CBT を実現するための工夫や仕掛けについて、その概要を記す。

【本項で取り上げる紙と鉛筆によるテスト「X 資格試験」】

- 必要なスキルセットとしては、「関連法規」「X 理論」の 2 つのサブ分野からなる「知識領域」と、「実技領域」からなる。
- 「知識領域」は、「関連法規」「X 理論」の各サブ分野について、各 40 問、計 80 問からなるテスト版を用いて測定を行う。いずれも多肢選択式の客観式テストで、サブ分野内における出題範囲が事前に受験者へ提示されている。
- 「実技領域」は、X 資格の実務で不可欠である「業務上のやり取り」を模した音声を聞き取り、その内容に関する問題項目に解答する。解答形式は記述式。音声は試験会場に設置されたスピーカーを用いて提示され、受験者はこの音声を聞きながら解答用紙に答案を記述する。
- 試験会場は全国 10 都市。年 3 回（4 月、8 月、12 月）実施。
- 受験資格はなし。誰でも受験できる。
- 時間割は以下の通り。
 - 10:00～12:00：「知識領域」（関連法規）
 - 13:00～15:00：「知識領域」（X 理論）
 - 15:30～16:30：「実技領域」
- 採点は、「知識領域」においてはマークシートリーダを用いて行われ、1 問 2 点、計 160 点満点で得点化する。「実技領域」については評価者が解答用紙に書かれた答案について、採点基準に照らして 40 点満点で採点する。
- 200 点満点のうち、150 点以上を合格とする。ただし、実技領域で 15 点以下であった受験者は、合計点で 150 点以上をとっていても不合格となる。
- 年 3 回行われるテストで使用されるテスト版は、作題会議で 1 年分の問題項目をまとめて作っておき、難易度が同じ程度となるように各テスト版に割り振って作成する。実技領域での音声も、同様に 1 年分をまとめて作成する。

年 3 回行われるテストで出題される項目については、問題作成委員の経験と勘により、その難易度が均一になるように作成されてきた。しかしながら、項目の難易度を事後的にしか検証できない仕組みであったため、出題された項目の困難度が難しい方に偏っていた回の受験者が不利になるのではないかといった意見が、外部からなされるようになった。そのため、事前に項目の難易度を検討しておく必要性が指摘された。あわせて、項目の難易度を表示するにあたり、これまで行われてきた正答数による方法ではなく、IRT を用いて標準化された困難度を用いることとされた。ただし「実技領域」の音声については、IRT による標準化された困難度を推定することが難しいことから、この例では多肢選択式の部分のみを分析対象とする。

IRT を用いた困難度・識別力の表示のためには、何らかの方法で正誤データを収集して分析にかけ手順が必須である。そのため、実際に合否判定のために受験する者とは別に、困難度・識別力を推定するための「モニター受験者」を用意し、本試験で出題される前の問題項目に解答させ、その正誤データを分析することで、困難度や識別力を推定する。本試験以外に行われるテストを「予備テスト」と呼ぶ。モニター受験者は X 資格試験に関係する団体を通じて 1000 名ほどを確保し、毎年行われる 3 回のテストに先立ち、予備テストを受験するように依頼する。

1 年目の実施スケジュールは、具体的に以下のとおりである（図 9）。

- 12 月：出題される項目候補の確定。
- 1 月：予備テストのためのテスト版を作成。知識領域の各サブ分野 160 問ずつ、計 320 問。必要となる項目数はサブ分野について $40 \text{ 問} \times 3 \text{ 回分} = 120 \text{ 項目}$ であるが、内容に不備のある項目や極端に難しい・やさしい項目が出ることを考慮し、40 項目を余分に含ませる。
- 2 月：予備テストの実施。「関連法規」と「X 理論」それぞれ 1 日かけて、320 問に解答させる。実施後、IRT による分析を行い、困難度と識別力を推定する。
- 3 月：予備テストの IRT 分析結果まとめ。320 問のうち、困難度が互いに似た値となるような項目を 3 つのテスト版に割り振り、40 問からなるテスト版を作成する。残った 40 項目のうち内容に不備がない項目を、次年度と困難度を比較するための「アンカー項目」として保存する。
- 4 月から 12 月：3 つのテスト版による試験を実施。

予備テストの結果見出された困難度は、その問題文とともに厳重に秘匿される。またモニター受験者に提示されたすべての項目は、テスト実施後に回収される。モニター受験者はテスト版を持ち帰ることも、問題を記録することも許されない。

次年度以降、同様にモニター受験者を用いて予備テストを 2 月に行い、困難度・識別力が均質化されたテスト版を用いてテストを行う。ただしこのままでは、毎年作成された項目が、年度をまたいで比較可能ではない能力尺度上で表示されることとなる。言い換えれば、ある年度のモニター受験者を用いて推定された困難度が、他のモニター受験者から得られた困難度と比較可能とはなっていないこととなる。このままでは、合否基準が実施年度によって統一されず、テストの公平性が損なわれるおそれがある。

年度をまたいで共通の能力尺度上で困難度を表示するために、次年度以降のモニター受験者に対しては、前年の予備テストで出題し、本試験で出題しなかったアンカー項目を混ぜて出題する。次年度以降においては、モニター受験者はサブ分野あたり 160 項目 + アンカー 40 項目 = 200 項目を解答する（図 10）。次年度以降の IRT 分析にあたっては、合否基準をそろえるために、その年度の IRT 分析での能力値尺度を前年度の能力値尺度に合わせる操作が必要となるが、この操作が「等化」にはかならない。

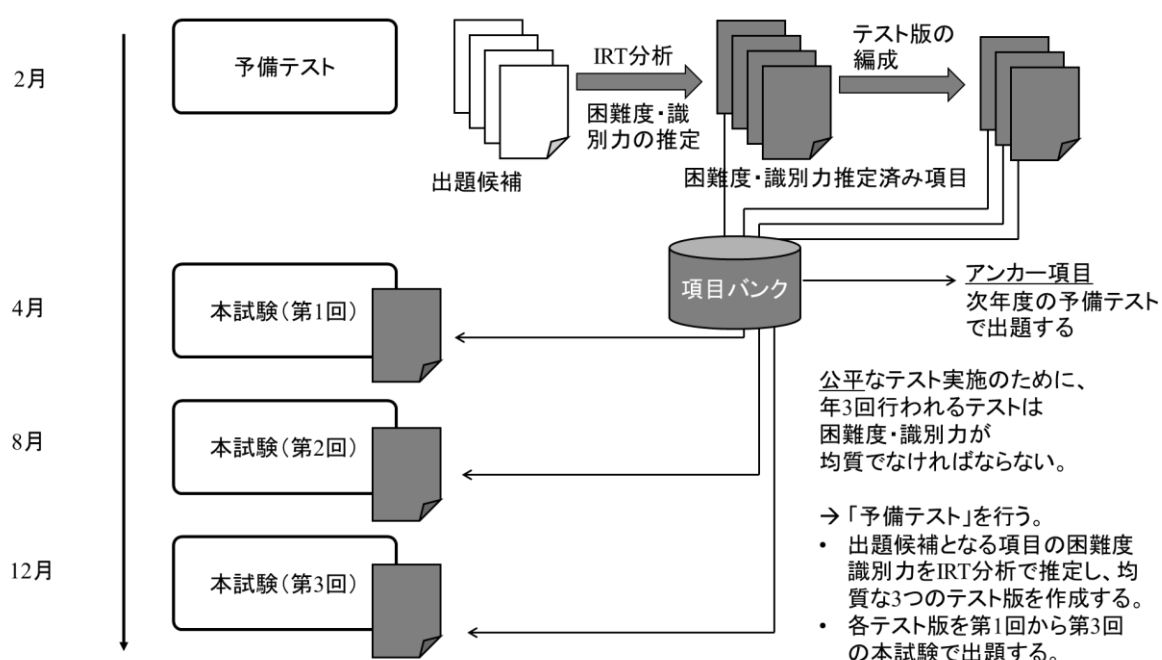


図9 本試験実施前に予備テストを行い、困難度・識別力がそろったテスト版を作成する手続き

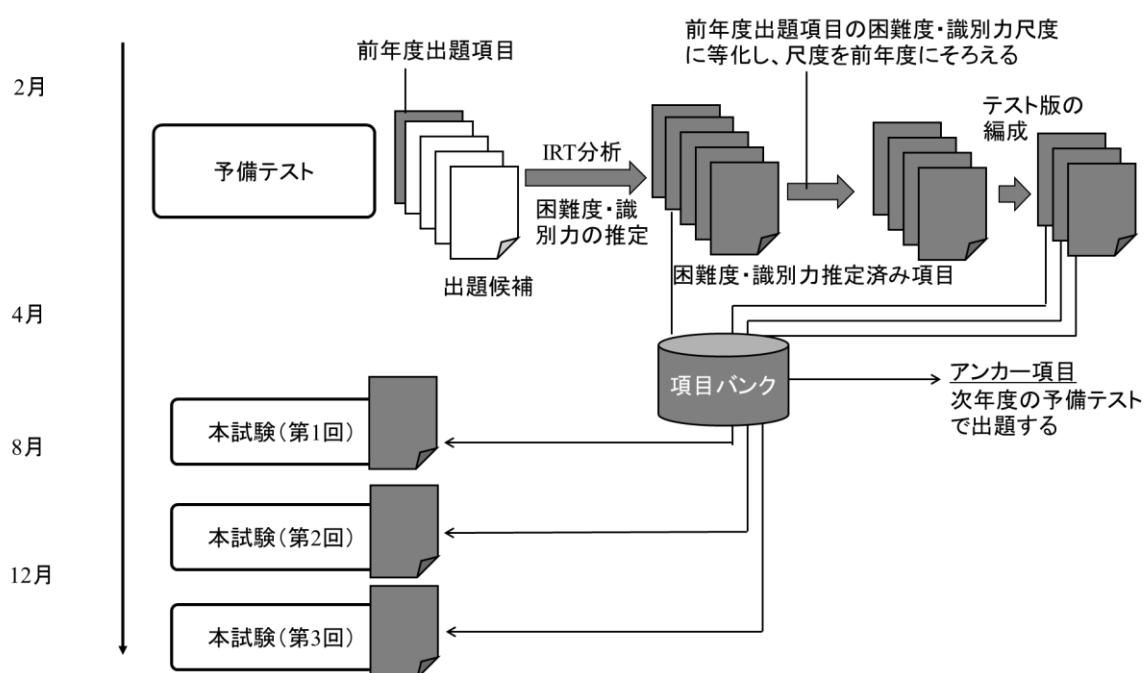


図10 本試験実施前に予備テストを行い、困難度・識別力がそろったテスト版を作成する手続き（2年目以降）

CBTによりテストを行うことは、ここで述べたIRTによる項目バンクを用いたテストの仕組みと親和性が高い。理由としては、異なる種類のテスト版を統一的に管理できること、紙と鉛筆の

テストでは複雑であったテスト冊子の編集作業を必要としないこと、出題履歴を受験者単位で記録するのが容易なことなどが挙げられる。

以上述べたテスト実施計画においては、毎年、1000名規模のモニター受験者を募集し、比較的多数の項目に解答することが求められる。1回あたりのテストで出題される項目数が多くなると、モニター受験者に提示される項目数も増加し、一人が解答できる項目数を超える可能性が出てくる。そのため、重複テスト分冊法⁵²などの手法を用いて、受験者一人あたりに解答させる項目数を減らすことを考えなければならない。

IRTを用いた標準化テストを行うための方法としては、上記に記したモニター受験者を毎年募集する手法が一般的である。一方で、本試験受験者に対して出題した問題項目の一部を、別の回の本試験で再出題する手法も提案されている。この場合は、本試験の問題冊子を受験者に持ち帰らせずに毎回回収し、受験者の「受験対策」のために一部の項目を公開するにとどめることとなる。

ただし、資格試験受験者の受験動向からみて、ある回で不合格となった受験者は、それ以降の回を連続して受験し、合格を目指すことがほとんどである。ある本試験で出題された項目が、ほかの本試験で再出題されると、過去の本試験と両方回答した受験者に対してのみ有利になることは明らかである。本試験の問題冊子を使いまわし、一度本試験受験者に対して出題した項目を別の本試験で再出題することは、公平なテストの実施を妨げるおそれがあるため、資格試験においては行われ⁵³ない。ただし、(受験者にとっての利便性は低下するが)同一年度に受験できるテストの回数を制限したり、一定期間再受験することを禁止したりすることで、項目の使い回しをしながら同一受験者が同一項目を目にしない仕組みを取り入れることも、技術的には可能であろう。

3. 2 公平なテストのために (1) 出題項目のコントロールと基準設定 (standard setting)

資格試験の公平性を高めるためには、出題された項目の困難度を統制することが求められる。もし、複数回行われるテストにおいて、ある回で出題されたテスト版だけが、困難度の低い項目ばかりであったならば、その回を受験した受験者だけが、項目がやさしかったために多くの正答を得ることができてしまう。正答数(または正答率)が一定以上であれば合格、という可否基準を設定していたならば、公平な可否判断とはならなくなるであろう。

多くの資格試験では、そのような事態を防ぐために、出題した項目の困難度が高かった場合に、事後的に可否判定基準を引き下げるといった補正の仕組みを導入している。しかし、事後的な操作による可否基準の操作が適切であったかどうかを検証することが難しいという欠点がある。

この欠点を克服するためには、モニター受験者(可否判定に得点を利用しない受験者)を対象

⁵² 重複テスト分冊法の実践例については、以下を参照のこと。

日本テスト学会[編](2010). 見直そう、テストを支える基本の技術と教育 金子書房, pp. 45-48.

また、重複テスト分冊法を用いてIRT分析を行い、多数の項目に対して項目特性値(困難度・識別力)を推定するための方法は、以下を参照のこと。

光永悠彦(2017). テストは何を測るのか—項目反応理論の考え方 ナカニシヤ出版, pp. 157-159.

⁵³ モニター受験者に出題した項目を本試験受験者に再出題することは、公平性を妨げない。ただし、モニター受験者が項目の内容を漏洩しないことが前提である。

に、これから出題されるであろう項目に対して解答を求める「予備テスト」を行い、困難度を IRT により推定しておき、推定された困難度が同じ程度になるようなテスト版を本試験実施前に用意する方法がある。

IRT による困難度指標は受験した者の能力分布に依存していないため、正答率のように、受験した者が誰なのかによって、数値の意味が変わるということがない。この特性を利用して、困難度指標を基準にすることで、「IRT による困難度が〇〇以上の項目に正答する確率が××パーセント以上である者を合格とする」という方法で合否判定のための基準を決めることができる（下記 2.の方法）。

能力値のカットオフポイントをモニター受験者のデータを用いて決めるためには、IRT を用いた尺度化を行っていることから、以下の2通りの方法のいずれかをとるのが適当である。

1. 予備テストにおいて、モニター受験者として「資格試験を受験し、合格した経験のある者」と「不合格あるいは未受験の者」をそれぞれ募集し、IRT による能力値の分布が前者と後者でどのように差がつくかを観測し、両者の境目に該当する能力値をカットオフポイントとする方法（対比グループ法）。
2. と同様の2群からなるモニター受験者を募集して予備テストを実施し、IRT による困難度を推定したうえで、複数の専門家に対して「合格者としてふさわしい難易度に相当する困難度の項目」を尋ね、その困難度に対応する能力値をカットオフポイントとする方法（Bookmark 法）。

前者の方法は、モニター受験者を規準集団と定め、本試験受験者の能力尺度を規準集団上に等化するテストの場合、直感的に合否基準を統制することが可能となる方法である。ただし、目標準拠型のテストである場合は、合格基準に相当する目標に未達の受験者であっても、測定誤差の影響で「合格」とされる能力値を超える可能性が出てくることが避けられない。より厳密な合否基準とするためには、後者の方法を併用し、項目の内容を加味した合否判断基準を設ければよい。ただし後者の手法においては、項目の困難度を判断するための専門家から、意見を募る必要が出てくる。

IRT による能力値を用いて受験者の得点を表示することとすれば、どの受験回を受験してもカットオフポイントをそろえることが可能となる。一方で、正答数による得点の表示をする場合に一般的であった「満点」や「0点」を基準とした得点の解釈をすることができない。出題された項目にすべて正答できれば「満点」であり、満点をとることを目安の一つとして受験対策をしてきた受験者がいたとすれば、IRT 化によってそのような基準がなくなることを意味する。

そのような事態を避けるためには、能力値の数値と、その意味するところを具体的に対応させ、得点と「その得点をとった者ができること」の大まかな目安を示せばよい。たとえば英語のテストで、200 点以上 300 点以下は「日常生活をある程度不自由なく送れる範囲の読解ができる」、300 点から 400 点は「日常生活を不便なく送ることができるような読解能力を有する」といったように、得点の範囲を「実際にできること」と対応させて表示することが行われている。これらの「できること」を記述した文言を「can-do ディスクリプタ」と呼び、can-do ディスクリプタと得点と

の対応表を「can-do リスト」と呼ぶ。

資格試験の場合も同様に、can-do リストを構築することにより、受験者のとった得点と「実際にできること」との対応関係を記述することができるようになり、受験者からみても「現在の学習の進捗状況」を把握することにつながり、より有用性の高いテストとなりうる。実際に can-do リストを構築するためには、受験者に対して「何ができるか」をアンケートする方法や、出題された項目内容と困難度を対応させる方法があるが、資格試験の場合は後者の方法をとることによって、出題された内容とリンクした形で「できること」を記述することにつながり、合否基準の設定にも役立つであろう。

3, 3 公平なテストのために（２）本人確認、受験履歴管理

公平なテストを実現するためには、受験者の受験環境を統制し、なるべく同一条件下で解答させることが必須である。同時に、テスト実施中は受験者同士が相談したり、試験室の外部から情報を収集したりしないように、受験者を監督する仕組みが必要である。あわせて、不正行為が行われた場合の対処法や、不測の事態（受験者の急病、天災による試験実施の障害など）への対応などについても、あらかじめマニュアル化しておく必要がある。

受験者が資格試験を受験するにあたり、以下のような３通りの形態が考えられる。

【集合型試験（PBT）】

- 受験者は、あらかじめ告知された受験会場に集合する。
- 受験者に対して、紙に印刷された問題冊子を提示し、解答を求める。
- 試験監督者が試験進行を管理する。また、受験者の本人確認も行う。
- 解答は解答用紙に記載される。解答用紙を収集し、採点する。

【集合型試験（CBT）】

- 受験者は、あらかじめ告知された受験会場に集合する。受験会場には、コンピュータ端末が受験者の人数分用意されている。
- 受験者に対して、コンピュータ端末を用いて問題を提示し、解答もコンピュータ端末に入力するよう求められる。
- 試験監督者が試験進行を管理する。また、受験者の本人確認も行う。
- 受験者が入力した解答はサーバに送信される。又は、受験会場の担当者が、コンピュータ端末に記録された解答データを、USB メモリ等を介して取り出し、収集する。採点は電子的に行われる。

【非集合型試験（IBT）】

- 受験者は、インターネット環境が整った任意の環境から受験できる。ただし、受験者が一人で部屋にすることが前提であり、受験中に他人との接触や通信機器の操作、メモの閲覧、その他これに疑われる一切の行為は禁止される。
- 受験者が項目に解答する直前に、受験環境のチェックが行われる。チェックの結果、公平な

テスト受験に適さないと判断された場合は、受験後であっても得点が無効となる。

- 受験者の所有するコンピュータ端末に、項目が配信され、解答もコンピュータ端末に入力するよう求められる。
- 受験者からみて試験監督者がいないように見えるが、実際には遠隔で試験監督者が受験者の挙動を監視している。試験監督者がリアルタイムで監視していない場合であっても、受験者の解答する様子がすべて録画されており、事後的に人間の目で検証する仕組みが備わっている。さらに、解答中の様子をとらえた映像はAI（人工知能）による画像解析にかけられ、パターンマッチングの結果「疑わしい挙動」と判断された場合には、警告メッセージが画面上に表示される。
- 受験者が入力した解答はサーバに送信され、電子的に採点が行われる。

年複数回実施されるテストの場合、出題される項目が、同一受験者に二度以上提示されると、そのような受験者だけが不当に正答しやすくなるため、公平なテストとはならなくなるおそれがある。そのため、受験者の受験履歴管理もまた重要である。あわせて、替え玉受験を防ぐために、顔写真による本人確認を行う仕組みも必要となる。本人確認は、受験会場に入室する際だけではなく、スコアレポートを返却する際にも行われることが多い。

上記を踏まえたうえで、テストをCBT化する場合について考える。公平性を最大限考慮したテストとするならば、自宅等の任意の場所で受験できるテストではなく、試験監督者がテストの進行をコントロールする【集合型試験（CBT）】とすることが望まれる。また、テストで出題される分野が2種類と少ないため、試験時間を紙筆テストの場合と同じように設定することが可能であり、受験者から見ても従来の紙筆テストから違和感を覚えることが少なくなるであろう⁵⁴。あわせて、全国各地にテストセンターを持つCBT事業者等に業務を委託することができれば、受験地の拡大を図ることも可能となる。

多肢選択式の「知識領域」については、問題項目の提示から解答の収集、採点までをすべてオンラインで行うことが可能であるが、記述式の「実技領域」については、コンピュータ端末へ受験者が記述した解答データを、評価者が評定することとなる。その際、電子的に採点を行う仕組みとして、評価者に専用端末を用意したうえで、テスト終了後に解答データを解答した個人が特定されない形で評価者の端末に表示させ、評価者から評定を収集するような「電子評定システム」を導入することも行われている。

IRTによる標準化テスト化を目指す場合は、これまでの資格試験で行われてきた作問体制に加えて、出題してきた項目の履歴を管理する仕組み（項目バンクの仕組み）と、項目バンクを管理

⁵⁴ 資格試験は、その制度上、ある回で不合格となった受験者が次回以降も連続して受験することがほとんどである。また、出題された項目（項目例のサンプルである場合を含む）の実物を民間のテスト対策業者が収集し、テストの受験動向の解説に使用することも多い。そのため、これまで紙筆式で行ってきたテストの「見た目」を大きく変えることは、受験者以外においても混乱を招くことが多い。このような混乱を最小限に抑えるためには、受験者に出題された項目を回収して秘匿するといった操作をせず、可能な限りの情報公開を行うことが必須であろう。一方で、受験者に対してすべての項目を開示せず、サンプル項目例のみを開示することが許されれば、共通尺度化（等化）を行うための計画もここに述べた以外の手法をとることができるようになり、より効率的なテストの実施が期待できる。

運用していくための組織・規程の構築が必要である。これらについては、専門のノウハウを持つ事業者に委託することで、テスト実施団体の事務的負担を減らすことができる。一方で、円滑なテストの実施にあたっては、テストの運営をするうえで専門的な知識（テスト理論への理解やテストの実務経験等の知見）をもつ者がテスト実施団体の側にも存在していることが望ましい。

4. テストの公平性とは

4. 1 ハイ・ステークスなテストと公平性

これまで、資格試験における公平性をどのように確保していくかについて、テストを実施する立場からコントロール可能な要因について触れてきた。すなわち、出題基準や問題内容の適切な設定にはじまり、難易度を適切に保つ工夫や、回をまたいで比較可能な得点を表示する方法といった点にわたるまで、「テストの質」を高めることで公平なテストを実現する手法について述べてきた。

しかしながら、資格試験から一步引いてみたとき、大学入試や就職試験といった「選抜目的のテスト」を実施する現場においてもまた、公平性の確保という問題が横たわっていることが見て取れる。選抜目的のテストでは、得点が高い受験者から順に合格・採用されるというテストの性格上、受験者にとって、不正行為によって不当にメリットを享受しようとする動機がはたらくしやすい。そしてその傾向は、テストが「難関」とされるほどに高まることが予想される。

資格試験の場合であっても、同様に「難関」とされるほど、不正行為を行うことで不当にもたらされるメリットが大きくなる。結果が受験者の一生を左右するようなテストである場合、そのテストは「ハイ・ステークスなテスト」と呼ばれ、制度設計上、公平性の確保により細心の注意が求められる。本節では、テストにおける公平性について述べるとともに、テスト制度全体において公平性をどのように確保していくかについて、記述する。

4. 2 公平なテストが求められる社会的背景

ところで、前項で述べた「公平性の確保に細心の注意が求められる」という表現の裏には、そのような注意を求めている主体の存在が仮定できる。一般的には、そのような主体は「社会全体」であり、資格試験や入試などのテスト制度そのものを求めている主体（資格を認定したいと考えている団体や、膨大な志願者のうちだれを合格させるかを決めたいと考えている大学など）だけでなく、そのテストを受験する者や、対策法を指南する者といった幅広い範囲の者が含まれるであろう。その意味で、公平性の議論は社会的な要素を含む。では、そのような議論は、どのような観点から行われるのであろうか。

資格試験ではないが、入試に求められる公平性のあり方を議論するアプローチとしては、以下のような観点が提案されている⁵⁵。

- 格差や機会平等といった社会的観点からの検討。たとえば、親の学歴や居住地域といった変数が、合格・不合格にどう影響しているかといった観点。
- テスト理論の立場から、計量心理学のアプローチによる検討。たとえば、どの入試科目を選択したかによって合格のしやすさが変わってしまうか否か、といった観点。
- 社会心理学の立場から行われる、「人間はこのようなテストの仕組みをより公平に感じる」といった検討。たとえば、学校推薦型選抜よりも学力試験による選抜の方がより公平だと感じやすい、といった観点。

前節まででは主に、テストを実施する立場から、主に「テスト理論」や「計量心理学」の観点において、公平性を検討し、高めようとする手法について述べてきた。しかし、テスト制度が置かれた社会的背景を抜きにしては、いかに公平なテストを実施できたとしても、そのテストが社会から受け入れられることはない。「資格試験が社会的に『公平』である、とみなされる」ためには、前提として、受験者や関係者（受験対策を指導する業者など）、あるいは資格を取るための養成課程に従事する者が、「資格試験の公平性が確かに高い」と感じられることが求められる。そのためには、テスト理論の立場からの検討を経て公平なテストを提供する以上の、さらなる取り組みが求められる。

4. 3 不公平だと疑われる可能性を排除するプロセス⁵⁶

社会心理学の立場から「テストが不公平ではないか」と疑念を抱かれる要素として、「受験環境の不均一性」が挙げられる。ある受験者だけが十分明るい環境下でテストを受験できる一方で、一部の受験者だけ電灯が点滅するような環境下で受験することを余儀なくされた場合、後者はそのパフォーマンスを十分発揮できず、得点が低くなってしまうことが予測される。後者の得点は、受験者本来のパフォーマンスを反映した値ではなくなり、結果的に公平性が阻害されたテストと

⁵⁵ これらの観点は、以下の文献を参考にした。

林 洋一郎・倉元直樹（2021）. 公正研究からみた大学入試 西郡 大[編] 大学入試の公平性・公正性 東北大学大学入試研究シリーズ 金子書房, pp. 2-27.

同文献では、社会心理学の観点からのアプローチに基づき、東北大学のA0入試（現・総合型選抜）を受験した者に対するアンケート結果から、受験者は主に手続き的公平性を重要ととらえる傾向があったことを報告している。志望動機や面接試験は透明性が低いととらえられる反面、学力試験はより公平性が高いととらえられている点が特筆される。

⁵⁶ 本項の記述は、以下の文献を参考にした。

西郡 大・倉元直樹（2021）. 大規模リスニングテストにおける「妥当性」と「均一性」 —IC プレーヤー試験体験に参加した高校生の意見分析— 西郡 大[編] 大学入試の公平性・公正性 東北大学大学入試研究シリーズ 金子書房, pp. 168-188.

同文献では、高校生に英語リスニングテストを体験受験させ、その感想をアンケート調査により尋ねた結果を分析しており、「社会心理学的な観点からの公平感」を高めるためには、試験環境に対する不安、特に、リスニングテストの機器の操作に対する不安を解消することが必要であると指摘している。

なってしまう。

2006 年度から、大学入試センター試験（現・大学入学共通テスト）において「英語」科目にリスニングが導入された。受験者は英文音声を聴いて設問に答えることが求められるが、万一、音声が届き取れない環境となった場合、その設問に対する解答ができなくなってしまう、公平なテストではなくなってしまうおそれが指摘された。教室にスピーカーを設置して、スピーカーから流れる音声を聴きとる方式のテストが検討されたが、受験者の座る机の位置によって音量が異なることがわかり、ひいては解答に影響するという研究結果がみられ、「大学入試センター試験のように全国一斉形式でリスニングテストを実施するためには、教室にスピーカーを備える方法ではなく、受験者個別に音源（IC プレーヤー）を配布して実施すべきである」という結論となった。

個別に音源を用意して、イヤホンから音声を聴取する方式によるテストは、IC プレーヤーの信頼性の向上と、IC プレーヤーが故障した際のトラブル対策として、公平性を損なわない形での救済策をどうするかが課題となった。これらの検討には時間を要したが、入念な検討の結果、今日に至るまで大学入学共通テストの英語リスニングテストは行われ続けている。

英語リスニングテストを導入する際には、公平性に対する配慮が最大限行われた。反面、実施マニュアルが膨大になり、試験監督者への負担が増えるといった事態も起こっている。万一の事態、特に受験環境が不均一になった（音声が出ない、途中で止まる、など）場合に備えることは、テスト制度に対する信頼を高めるために必要な事項であるが、テストの実施可能性（フィージビリティ）とどのように両立するかが課題となるであろう。

4. 4 測定の専門家が考える「テスト規準」にのっとった公平性

本章の多くの部分で、テストの公平性を高めるために、「テストの質」のコントロールが重要であることを指摘した。この考え方は、テストの仕組みを開発する立場の研究者や実践家が考えている「テストのあるべき形」から出発した「テスト規準」において、もっとも重要な事項の一つととらえられている要素である。

日本における代表的な「テスト規準」に、日本テスト学会テスト規準作成委員会によって編集された「テスト・スタンダード」がある⁵⁷。ここでは「テストの公平性」について、テストを実施する側の責務として、以下のように記述されている。

受験者は、テストのすべての過程において年齢、性、国籍、障害の有無などによって差別されてはならない。また、質問項目の表現は、特定集団の成員に不快感を生じさせないように配慮されなければならない⁵⁸。

また「テスト・スタンダード」において、テストを実施する側が公平性を考えなければならない具体的な要素として、

⁵⁷ 日本テスト学会[編] (2007). テスト・スタンダード 日本テストの将来に向けて 金子書房

⁵⁸ 前掲書 p.138.

- 受験資格における公平性（受験資格を設ける場合は、要件を明確化し、受験要項などに記載して周知しなければならない、など）
- 実施方法の違いへの配慮（特定の受験者にとって回答方法による不利が生じないようにする、など）
- 質問項目の表現上の配慮（問題の表現に差別的なものや無用な誤解を生じさせるものが含まれないようにする、など）
- 結果の取り扱い（得点の算出方法や結果の利用法などについて、受験者が疑念を抱くことのないように、テストの利用目的に支障のない範囲でテスト情報の開示を検討する）

といった点が挙げられている⁵⁹。

これらの点については、いずれも「テストを行う側」がコントロールできる範囲で留意すべきことがらであり、いずれも「テストの質」を高める範疇に含まれる要素である。テスト・スタンダードは、テストを開発する側が、最低限考慮しなければならない注意点を記したものであり、その内容は計量心理学者などの専門家により吟味したものである⁶⁰。公平なテストのあり方を考える上で、いずれも重要な視点を含んでいるといえる。

4. 5 「テスト・スタンダード」における CBT

「テスト・スタンダード」では、CBT のあり方についても記載があるが、CBT の公平性に直接関係する要素としては、以下のようなものが挙げられる⁶¹。

- コンピュータやテスト実施環境の仕様を適切に定め、明示すること。またその仕様に沿った実施環境を用意すること。
- ユーザインタフェースを工夫すること。特にユーザビリティ、アクセシビリティ、ユニバーサルデザインに留意すること。また、例題や練習問題を用意し、受験者の誤操作を防ぐ工夫をすること。
- 問題や解答データ、結果データの機密保持ができること。また、本人確認の仕組みを含め、テストシステム全体が一貫したセキュリティポリシーの下で構築されていること。
- 不測の事態に対応できること。障害が発生した場合でも速やかにテストが再開できること。
- CBT とそれ以外の形式のテストが併用されている場合、得点の比較可能性を検討し、得点の換算表を用意すること。得点が比較可能ではない場合は、別のテストであることを明確にすること。

⁵⁹ 前掲書 pp.138-141.

⁶⁰ 公平性を第三者によって審査するためのガイドラインを設け、それに沿って公平性を評価するという考え方もある。北米で行われるテストの場合のガイドラインとしては、たとえば Educational Testing Service (ETS) による ETS Guidelines for Developing Fair Tests and Communications (2022) があるが、適用にあたってはテストが実施される文化圏の違いを考慮すべきであろう。

<https://www.ets.org/content/dam/ets-org/pdfs/about/fair-tests-and-communications.pdf>

⁶¹ 前掲書 pp.118-133.

CBT を公平なものとするためには、上記のような幅広い点に留意しなければならない。これらの技術的要件を満たすためには、コンピュータ技術者やネットワーク技術者、計量心理学者、テストの企画・作問及び設計実施管理者（テストディベロッパー）といった広範囲にわたる専門家の知見を集積させることが求められる。また CBT 化によって新たに解決しなければならない点があるため、CBT を初めて導入するテスト事業者にとって研究的要素を含むことに注意が必要である。

4. 6 公平なテストを実施するための「マニュアル化」

CBT ではない形式のテスト、たとえば紙と鉛筆によるテストは、実施上の長い経験を有し、公平性を保つための処置（たとえば、事後的に問題不成立が発覚した場合の対応など）も多く行われ、受験者や関係者からみて「不公平ではないか」という指摘がなされるべくないよう工夫が重ねられてきた。同時にこれらの経験は、「実施マニュアル」という形でそのノウハウが明文化され、テストを実施する主体が受け継いできた。

紙と鉛筆のテストにおいても、マークシートを導入することにより、新たに公平性への配慮が求められる場面が出てきていた（例：マーク内容の訂正は消しゴムで完全にマークを消すことが求められているが、消し方が不十分である場合、マークが訂正されたかどうかを誰がどのように判断するのか）。このような場面への対処法は、マニュアルに明文化され、そのルールに沿った形でテストを運用することで、公平なテストになるように工夫されてきた。

CBT 導入に際してもこのようなマニュアル化は必須であり、受験者に対する適切な説明のためにもノウハウの蓄積は必要である。さらに、マニュアルは適宜アップデートされなければならない。

ここで問題となるのは、CBT の実施の一部を外部に委託した場合に、CBT 委託先の業者が持っている CBT 実施のノウハウと、テスト実施する団体が（CBT 化以前に確立した）実施のノウハウの間に齟齬があった場合である。CBT の経験が長い委託業者が確立した実施マニュアル策定の指針を理解し、必要に応じてテスト実施団体の目指すテストの実施手法を取り入れる（カスタマイズする）ように要求するためには、CBT の仕組みで何ができ、何ができないのかについて、テスト実施団体の側にも一定の理解がなければならないであろう。

4. 7 公平な CBT 実施のために必要な人的リソースと経験知

CBT を実施するテスト実施団体は、自らが行うテストが公平であることを外部に説明するために、受験者に対する「テスト実施要項」による説明とは別に、模擬テストを体験させる場を提供することが一般的である。

一般に、CBT は多くの受験者にとってなじみがなく、実際のテストで何が行われるのかについてのイメージに乏しい。また、普段コンピュータを扱う上でなじみのある操作が、不正行為とみ

なされる場合もありうる⁶²。CBT では受験者にとって、解答操作上の禁止事項が多くなりがちであり、これらを説明するためには文書による理解促進だけではなく、動画などを用いた直感的に理解されやすい説明が必要である。

また、受験者に「お試し受験」の機会を与えることは、受験者の CBT 実施に対する不安の低減をもたらし、受験者本来のパフォーマンスを発揮させることにもつながる。さらに、試験実施会場における CBT 受験上のトラブルについては、専門のトレーニングを受けた者がサポートすることで、公平性を保ったままトラブルに対応できるようになるが、トラブルで動揺している受験者に安心感を与えつつトラブルに対応するためには、試験監督者に相応のスキルが求められる。

受験者が不安に思うことなく、余分なプレッシャーを感じずにテストを受験できるようになるためには、ここで述べた方法以外にも多くの実施上の工夫が求められる。テスト受験場面はもともと受験者を不安にさせる要因となりうるが、その不安を高めないようにする努力をすることで、受験者の不公平感を減らすことが期待できる。多くの心理学の研究で、人間が不安に陥る要因や不安を低減するための方法が検討されている。

以上より、CBT の公平な実施のためには、コンピュータ端末を導入することにとどまらず、熟練した試験監督者や、人間の行動について知見をもつ心理学者の助言が必要であるといえる⁶³。これらの専門的知見については、CBT の実施を請け負う側が一定のノウハウを有していることが多いが、実施団体の側にもそのようなノウハウを共有する環境があることが望ましい。

⁶² たとえば、テスト中にテスト解答アプリ以外の他のアプリケーションの操作をしようとする、カメラを OFF にする、画面以外に視線をずらす、などが挙げられる。

⁶³ CBT の公平な実施に関する先行事例を研究することにより、公平性を高めることにもつながるであろうが、それらを研究する場（学会・研究会等）に参加することによって、さらに効率的な公平性の向上が見込まれるであろう。

付録書 B

アンケート調査票

国家試験等の実施状況に関するアンケート（実施：高圧ガス保安協会）

このアンケートについて

【目 的】

本アンケートでは、資格試験（国家試験等を含む）に関する実施の方式や状況等について把握することを目的としています。是非ともご協力をお願い致します。

【対 象】

資格試験（国家試験等を含む）を実施している機関・団体・学校様

【回答締切期日】

2023 年 1 月 13 日（金）

【用語の説明】

本アンケートでは、資格試験の実施方式について以下のとおり定義します。

試験実施方式の定義

実施方式		特徴
PBT 方式		方法： <u>紙面（問題・解答用紙等）を配布・回収する方法</u> 場所： <u>指定された会場</u> 時間： <u>全国一同に実施</u>
非 PBT 方式	CBT 方式	方法： <u>コンピュータ端末（会場に設置のもの）を使用する方法</u> 場所： <u>指定された会場（テストセンター）</u> 時間： <u>①個別の時間帯で試験を実施</u> 又は <u>②全国一斉に実施</u>
	IBT 方式	方法： <u>コンピュータ端末（任意のもの）を使用する方法</u> 場所： <u>任意の場所</u> 時間： <u>①個別の時間帯で試験を実施</u> 又は <u>②全国一斉に実施</u>
その他		<u>上記に分類されない方法</u> 例： <u>自宅学習型（郵便制度を利用し、解答、レポート提出等を行わせる方法）等</u>

本アンケートはインターネット上からでもご回答頂けます。
下記の 2 次元バーコード又は URL からアクセスしてください。

2 次元
バーコード



URL

<https://onl.la/ddqP6Yj>

貴機関・団体・学校等について

貴機関・団体・学校及び入力担当者様についてご記入ください。

1. 貴機関・団体・学校について

貴機関・団体・学校名	
所在地	〒

	(複数回答可)	☐ C. PBT 方式の試験を中心に実施予定 ☐ D. その他 ()
--	---------	---

【2】試験名称 (記述式)		
試験の実施状況	問題の公開 (複数回答可)	☐ A. 常に HP 等で公開している ☐ B. 一定期間のみ HP 等で公開している ☐ C. 過去問題集を発刊している ☐ D. 公開していない ☐ E. その他 ()
	本人確認の為の 事前提出物 (複数回答可)	☐ A. 受験票 (写真有) ☐ B. 受験票 (写真無) ☐ C. 公的身分証 (写真有) ☐ D. 公的身分証 (写真無) ☐ E. 免状や資格証 (写し) ☐ F. 特になし ☐ G. その他 ()
	直近の年間受験者数 (記述式)	約 名／年
	問題の解答形式 (複数回答可)	☐ A. 選択式 (択一・択多) ☐ B. 記述 ☐ C. 口述 ☐ D. 実技 ☐ E. その他 ()
	現在の試験実施方式 (複数回答可)	☐ A. PBT 方式 ☐ B. CBT 方式 ☐ C. IBT 方式 ☐ D. その他 ()
	今後の試験実施計画 (複数回答可)	☐ A. 非 PBT 方式の試験を中心に実施予定 ☐ B. PBT 方式と非 PBT 方式の試験の両方を実施予定 ☐ C. PBT 方式の試験を中心に実施予定 ☐ D. その他 ()

非 PBT 方式の試験実施について

非 PBT 方式の試験を採用している場合 ➡ このまま質問 4 へお進みください。
採用していない場合 ➡ 質問 9 へお進みください。

4. 非 PBT 方式を採用している場合、その試験システムについてあてはまるものを次の A～C から 1つお選びください。複数の試験システムを導入している場合は、代表的なもの1つについてお選びください。

☐ A. 自組織の専用システムを構築

☐ B. 外部の事業者が一般向けに提供するサービスを利用

└─▶<B. を選択された場合ご記入ください>事業者名：_____

サービス名：_____

☐ C. その他 _____

5. 非 PBT 方式を採用している場合、試験会場の運営方法についてあてはまるものを次の A～C から 1つお選びください。複数の運営方法がある場合には、代表的なもの1つについてお選びください。

☐ A. 自組織にて直接運営

☐ B. 外部の事業者に運営を委託

└─▶<B. を選択された場合ご記入ください>事業者名：_____

サービス名：_____

☐ C. その他 _____

6. 非 PBT 方式を採用している場合、その理由として最もあてはまるものを次の A～F から 1つお選びください。

☐ A. 受験者の実施利便性の向上

☐ B. 事務局の労力の低減

☐ C. 受験機会の拡大

☐ D. 問題の再利用が可能

☐ E. 感染症対策

☐ F. その他 _____

7. 非 PBT 方式を採用している場合、その導入時の課題として最もあてはまるものを次の A～F から 1つお選びください。

☐ A. 導入コストが高額

☐ B. 問題作成数が不足

☐ C. 合否判定上の課題

☐ D. 本人確認が困難

☐ E. システムの煩雑化

☐ F. その他 _____

8. 非 PBT 方式を採用した後、その効果として最もあてはまるものを次の A～F から 1つお選びください。

☐ A. 受験者の実施利便性の向上 ☐ B. 事務局の労力の低減 ☐ C. 受験機会の拡大

☐ D. 問題の再利用が可能 ☐ E. 感染症対策 ☐ F. その他 _____

9. 非 PBT 方式を採用していない場合、その理由として最もあてはまるものを次の A～F から 1 つお選びください。

☐ A. 導入コストが高額 ☐ B. 問題作成数が不足 ☐ C. 合否判定上の課題

☐ D. 本人確認が困難 ☐ E. システムの煩雑化 ☐ F. その他 _____

10. 非 PBT 方式の試験について、お寄せいただける情報がございましたらご記入ください。

非 PBT 方式の試験に関するヒアリングのお願い

11. 非 PBT 方式での試験の実施状況（例：試験運営の方針や課題、本人確認の方法等）について、貴機関・団体・学校にヒアリング（リモート・1 時間程度）をさせていただいても宜しいでしょうか。ヒアリングの実施時期は、2023 年 2 月頃を予定しています。ヒアリングをお受けいただける場合、上記 2. にご記載の入力担当者様に事前にご連絡をいたします。

☐ A. ヒアリングを受けられます（可）。

☐ B. ヒアリングは受けられません（否）。

ご協力ありがとうございました。